De novo DNA CYTOSINE METHYLTRANSFERASE GENES, POLYPEPTIDES AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U.S. Application No. 09/720,086, which is the National Stage of International Application No. PCT/US99/14373, filed June 25, 1999 and published in English under PCT Article 21(2)), which claims the benefit of U.S. Application No. 60/093,993, filed July 24, 1998, and U.S. Application No. 60/090,906, filed June 25, 1998. The content of all the aforesaid applications are relied upon and incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

Field of the Invention

[0002] The present invention relates generally to the fields of molecular biology, developmental biology, cancer biology and medical therapeutics. Specifically, the present invention relates to novel *de novo* DNA cytosine methyltransferases. More specifically, isolated nucleic acid molecules are provided encoding mouse Dnmt3a, and Dnmt3b and human DNMT3A and DNMT3B *de novo* DNA cytosine methyltransferase genes. Dnmt3a and Dnmt3b mouse and DNMT3A and DNMT3B human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are isolated nucleic acid molecules encoding mouse Dnmt3a2 and human DNMT3A2, which are small forms of the corresponding Dnmt3a mouse and DNMT3A human genes. Dnmt3a2 mouse and DNMT3A2 human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. The invention further relates to an *in vitro* method for cytosine C5 methylation. Also provided is a diagnostic method for neoplastic disorders, and methods of gene therapy using the polynucleotides of the invention.

Related Art

[0003] Methylation at the C-5 position of cytosine predominantly in CpG dinucleotides is the major form of DNA modification in vertebrate and invertebrate animals, plants, and fungi. Two distinctive enzymatic activities have been shown to be present in these organisms. The *de novo* DNA cytosine methyltransferase, whose expression is tightly regulated in development, methylates unmodified CpG sites to establish tissue or gene-specific methylation patterns. The maintenance methyltransferase transfers a methyl group to cytosine in hemi-methylated CpG sites in newly replicated DNA, thus functioning to maintain clonal inheritance of the existing methylation patterns.

[0004] *De novo* methylation of genomic DNA is a developmentally regulated process (Jähaner, D. and Jaenish, R., "DNA Methylation in Early Mammalian Development," In *DNA Methylation: Biochemistry and Biological Significance*, Razin, A. *et al., eds.,* Springer-Verlag (1984) pp. 189-219 and Razin, A., and Cedar, H., "DNA Methylation and Embryogenesis," in *DNA Methylation: Molecular Biology and Biological Significance*, Jost., J. P. *et al., eds.,* Birkhäuser Verlag, Basel, Switzerland (1993) pp. 343-357). It plays a pivotal role in the establishment of parental-specific methylation patterns of imprinted genes (Chaillet, J. R. *et al., Cell 66*:77-83 (1991); Stöger, R. *et al., Cell* 73:61-71 (1993); Brandeis, M. *et al., EMBO J. 12*:3669-3677 (1993); Tremblay, K. D. *et al., Nature Genet. 9*:407-413 (1995); and Tucker, K. L. *et al., Genes Dev. 10*:1008-1020 (1996)), and in the regulation of X chromosome inactivation in mammals (Brockdoff, N. "Convergent Themes in X Chromosome Inactivation and Autosomal Imprinting," in *Genomic Imprinting: Frontiers in Molecular Biology*, Reik, W. and Sorani, A. *eds.,* IRL Press Oxford (1997) pp. 191-210; Ariel, M. *et al., Nature Genet. 9*:312-315 (1995); and Zucotti, M. and Monk, M. *Nature Genet. 9*:316-320 (1995)).

[0005] Thus, C5 methylation is a tightly regulated biological process important in the control of gene regulation. Additionally, aberrant *de novo* methylation can

lead to undesirable consequences. For example, *de novo* methylation of growth regulatory genes in somatic tissues is associated with tumorigenesis in humans (Laird, P. W. and Jaenisch, R. *Ann. Rev. Genet. 30*:441-464 (1996); Baylin, S. B. *et al., Adv. Cancer. Res. 72*:141-196 (1998); and Jones, P. A. and Gonzalgo, M. L. *Proc. Natl. Acad. Sci. USA 94*:2103-2105 (1997)).

[0006] The gene encoding the major maintenance methyltransferase, *Dnmt*1, was first cloned in mice (Bestor, T. H. *et al., J. Mol. Biol. 203*:971-983 (1988), and the homologous genes were subsequently cloned from a number of organisms, including *Arabidoposis*, sea urchin, chick, and human. *Dnmt*1 is expressed ubiquitously in human and mouse tissues. Targeted disruption of *Dnmt*1 results in a genome-wide loss of cytosine methylation and embryonic lethality (Li *et al.*, 1992). Interestingly, *Dnmt*1 is dispensable for the survival and growth of the embryonic stem cells, but appears to be required for the proliferation of differentiated somatic cells (Lei *et al.*,, 1996). Although it has been shown that the enzyme encoded by *Dnmt*1 can methylate DNA *de novo in vitro* (Bestor, 1992), there is no evidence that *Dnmt*1 is directly involved in *de novo* methylation in normal development. *Dnmt*1 appears to function primarily as a maintenance methyltransferase because of its strong preference for hemi-methylated DNA and direct association with newly replicated DNA (Leonhardt, H. *et al.*,, *Cell 71*:865-873 (1992)). Additionally, ES cells homozygous for a null mutation of *Dnmt*1 can methylate newly integrated retroviral DNA, suggesting that *Dnmt*1 is not required for *de novo* methylation and an independently encoded *de novo* DNA cytosine methyltransferase is present in mammalian cells (Lei *et al.*,, 1996).

[0007] Various methods of disrupting *Dnmt*1 protein activity are known to those skilled in the art. For example, see PCT Publication No. WO92/06985, wherein mechanism based inhibitors are discussed. Applications involving antisense technology are also known; U.S. Patent No. 5578716 discloses the use of antisense oligonucleotides to inhibit *Dnmt*1 activity, and Szyf *et al.*,, *J. Biol.*

Chem. 267: 12831-12836, 1992, demonstrates that myogenic differentiation can be affected through the antisense inhibition of *Dnmt*1 protein activity.

[0008] Thus, while there is a significant amount of knowledge in the art regarding the maintenance C5 methyltransferase (*Dnmt*1), there is no information regarding nucleic acid or protein structure and expression or enzymatic properties of the *de novo* C5 methyltransferase in mammals.

SUMMARY OF THE INVENTION

[0009] A first aspect of the invention provides novel *de novo* DNA cytosine methyltransferase nucleic acids and polypeptides that are not available in the art.

[0010] More specifically, isolated nucleic acid molecules are provided encoding mouse Dnmt3a, and Dnmt3b and human DNMT3A and DNMT3B *de novo* DNA cytosine methyltransferase genes. Dnmt3a and Dnmt3b mouse and DNMT3A and DNMT3B human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are isolated nucleic acid molecules encoding mouse Dnmt3a2 and human DNMT3A2, which are small forms of the corresponding Dnmt3a mouse and DNMT3A human genes. Dnmt3a2 mouse and DNMT3A2 human polypeptides are also provided, as are vectors, host cells and recombinant methods for producing the same. Also provided are Dnmt3a2 mouse and human DNMT3A2 promoter sequences.

[0011] A second aspect of the invention relates to *de novo* DNA cytosine methyltransferase recombinant materials and methods for their production.

[0012] A third aspect of the invention relates to the production of recombinant *de novo* DNA cytosine methyltransferase polypeptides.

[0013] A fourth aspect of the invention relates to methods for using such *de novo* DNA cytosine methyltransferase polypeptides and polynucleotides. Such uses include the treatment of neoplastic disorders, among others.

[0014] Yet another aspect of the invention relates to diagnostic assays for the detection of diseases associated with inappropriate *de novo* DNA cytosine

methyltransferase activity or levels and mutations in *de novo* DNA cytosine methyltransferases that might lead to neoplastic disorders.

BRIEF DESCRIPTION OF THE FIGURES

[0015] Figures 1A-1D shows the nucleotide sequences of mouse Dnmt3a and Dnmt3b and human DNMT3A and DNMT3B genes respectively.

[0016] Figures 2A-2D shows the deduced amino acid sequence of mouse Dnmt3a and Dnmt3b and human DNMT3A and DNMT3B genes, respectively. Sequences are presented in single letter amino acid code.

[0017] Figure 3A shows a comparison of mouse Dnmt3a and Dnmt3b amino acid sequences, and Figure 3B presents a comparison of the protein sequences of human DNMT3A and DNMT3B1.

[0018] Figure 4A presents a schematic comparison of mouse *Dnmt1, Dnmt2,* Dnmt3a and Dnmt3b protein structures. Figure 4B presents a schematic of the DNMT3A, DNMT3B and zebrafish Zmt3 proteins. Figure 4C and 4D present a schematic of the human DNMT3B gene organization and exon/intron junction sequences.

[0019] Figure 5A presents a comparison of highly conserved protein structural motifs for eukaryotic and prokaryotic C5 methyltransferase. Figure 5B presents a sequence alignment of the C-rich domain of vertebrate DNMT3 proteins and the X-lined ATRX gene. Figure 5C presents a non-rooted phylogenic tree of methyltransferase proteins.

[0020] Figures 6A-6C demonstrates the expression of Dnmt3a and Dnmt3b in mouse adult tissues, embryos, and ES cells by northern blot.

[0021] Figures 7A-7D demonstrates *in vitro* methyltransferase activities of mouse Dnmt3a and Dnmt3b proteins.

[0022] Figure 8 demonstrates *in vitro* analysis of *de novo* and maintenance activities of Dnmt3a, Dnmt3b1 and Dnmt3b2 proteins.

[0023] Figure 9 presents Northern blot expression analysis of DNMT3A and DNMT3B.

[0024] Figure 10 presents DNMT3 Northern Blot expression analysis of DNMT3A and DNMT3B in human tumor cell lines.

[0025] Figures 11A-11F present the identification of novel isoforms of Dnmt3a and Dnmt3b proteins. Figure 11A shows a schematic diagram of Dnmt3a and Dnmt3b proteins. The conserved PWWP and PHD domains, the methyltransferase motifs (I, IV, VI, IX, and X), and the sites of alternative splicing are indicated (the C-terminal 45 amino acids of Dnmt3b5 are out of frame and shown as an open bar). The locations of the epitopes for the Dnmt3 antibodies (164, 157, and 64B1446) are also shown. Figure 11B demonstrates the specificity of the Dnmt3a and Dnmt3b antibodies. Mouse (m) and human (h) Dnmt3a and Dnmt3b were expressed as GFP fusion proteins in Cos-7 cells and analyzed by immunoblotting with the indicated antibodies. Figure 11C demonstrates that ES cells express Dnmt3b1 and Dnmt3b6. Cell lysates from wt (J1), Dnmt3a$^{-/-}$ (6aa), Dnmt3b$^{-/-}$ (8bb), and [Dnmt3a$^{-/-}$, Dnmt3b$^{-/-}$] double mutant (7aabb) ES cells as well as Cos-7 cells transfected with different Dnmt3b isoforms were immunoblotted with Dnmt3b-specific antibody 157. Figure 11D demonstrates that ES cells express at least two forms of Dnmt3a proteins, Dnmt3a and Dnmt3a2. The same ES cell lysates as described in Figure 11C as well as control Dnmt3a protein expressed in Cos-7 cells were immunoblotted with Dnmt3a-specific antibody 164 (lanes 1-5) and the mAb 64B1446 (lanes 6-10). Figure 11E demonstrates that Dnmt3a2 co-migrates with a truncated Dnmt3a protein lacking the N-terminal 219 amino acid residues. Plasmid constructs encoding N-terminally truncated Dnmt3a proteins or vector alone were transfected into 6aa ES cells. The overexpressed proteins as well as endogenous Dnmt3a2 (from J1 cells) were immunoprecipitated and detected with antibody 64B1446. Note that lysis buffer containing low salt (150 mM NaCl) could not extract Dnmt3a and Dnmt3b1. Figure 11F illustrates that Dnmt3a2 cannot be derived from *Dnmt3a* cDNA. Plasmid construct encoding Dnmt3a or vector alone

was transfected into 6aa ES cells. The transfected cells as well as J1 cells were lysed and immunoblotted with antibody 64B1446.

[0026] Figures 12A-12C demonstrate that Dnmt3a and Dnmt3a2 are encoded by distinct transcripts. Figure 12A presents the structure of mouse and human *Dnmt3a* gene, mRNAs and proteins. Exons are shown as black bars. The *Dnmt3a2* unique exons are indicated by "*". Dnmt3a and Dnmt3a2 proteins have identical amino acid sequences except that Dnmt3a has 219 (mouse) or 223 (human) extra residues at the N terminus (human DNMT3A amino acid numbering is shown in parenthesis). The primers used for RT-PCR are shown under the corresponding exons (F, forward; R, reverse). The probes (lines under the Dnmt3a protein) that are used for Northern hybridization represent the corresponding cDNA fragments. Figure 12B presents Northern blots of total RNA (20 μg per lane) from NIH 3T3, J1, and 6aa cells were probed with Probe 1 (lanes 1-3) or Probe 2 (lanes 4-6). As a loading control, ethidium bromide (EB) staining of 28S rRNA was shown (lanes 7-9). Figure 12C presents RT-PCR results of Dnmt3a expression. Total RNA from J1 cells was reverse transcribed using poly $(dT)_{12-18}$ and the resulting cDNAs were subjected to PCR amplification with the indicated Dnmt3a primers. Dnmt3a cDNA was used as a positive control.

[0027] Figures 13A-13F present the nucleotide and predicted amino acid sequences of mouse Dnmt3a2 and human DNMT3A2. Figure 13A presents mouse *Dnmt3a2* cDNA sequence. Nucleotides 148-2217 represent coding sequence. Figure 13B presents mouse Dnmt3a2 predicted amino acid sequence. Figure 13C presents human *DNMT3A2* cDNA sequence. Nucleotides 217-2286 represent coding sequence. Figure 13D presents human DNMT3A2 predicted amino acid sequence. Figures 13E1-E4 present an alignment of the human *DNMT3A2* and mouse *Dnmt3a2* cDNA sequences. Figure 13F presents an alignment of the human DNMT3A2 and mouse Dnmt3a2 predicted amino acid sequences.

[0028] Figures 14A-14B demonstrate that a region 5' adjacent to the *Dnmt3a2* unique exon has promoter activity. Figure 14A presents a schematic representation of the luciferase reporter constructs. The genomic region that contains the *Dnmt3a2* unique exon (exon 7, black bar) embedded in a GC-rich region (striped bar) is shown at the top. The putative Dnmt3a2 transcription start site is indicated. In the reporter constructs, a 2.0-kb genomic fragment that contains part of exon 7 and the putative promoter region was inserted in both orientations upstream of the cDNA encoding the firefly luciferase (luc) followed by the SV40 late poly(A) signal (pA). Figure 14B demonstrates a luciferase activity assay. ES cells and NIH 3T3 cells were transfected with the reporter constructs (P2-luc and P2R-luc) and the empty vector pGL-3-Basic (luc) in the presence of pRL-TK (expresses *Renilla* luciferase), and luciferase activities were measured by luminescence. Firefly luciferase activity was normalized to *Renilla* luciferase activity to minimize transfection efficiency variations. The results were expressed as relative activity using the background activity generated by the empty vector as baseline. Each bar represents the mean + standard deviation of data from six independent reactions performed in two separate experiments.

[0029] Figures 15A-15D demonstrate that deletion of the putative *Dnmt3a2* promoter region abolishes *Dnmt3a2* transcripts and Dnmt3a2 protein. Figure 15A illustrates the targeted disruption of *Dnmt3a*2. The wild type genomic DNA structure with exons (black bars) and a GC-rich region (striped bar) in the putative *Dnmt3a2* promoter region is shown at the top. The putative transcription and translation start sites for Dnmt3a2 are indicated. In the P2 targeting vector, a 2.1-kb genomic fragment encompassing the *Dnmt3a2* unique exon and the putative promoter region was replaced with an hCMV-hygTK cassette in an opposite transcriptional orientation as *Dnmt3*a. A PGK-DTA cassette was introduced for negative selection to increase the targeting frequency. The location of the probe for Southern hybridization and *Sca* I (S) sites are also shown. Figure 15B presents Southern analysis of the genotype of ES cell lines. Genomic DNA was digested with *Sca* I and hybridized with the indicated probe. The 17 kb

untargeted allele (wt/6aa) and the 9 kb targeted allele (P2) are indicated. Figure 15C presents Northern analysis of total RNA from the ES cell lines. Note the intensity of the 4.0 kb and 4.2 kb bands was reduced by half in *Dnmt3a*$^{+/-}$ cells and was diminished in 296 cells. The 28S rRNA stained with ethidium bromide is shown as a loading control (bottom panel). Figure 15D presents immunoprecipitation and immunoblotting analyses of the ES cell lines with antibody 64B1446.

[0030] Figures 16A-16D demonstrate that Dnmt3a and Dnmt3a2 have similar methyltransferase activity but exhibit different subcellular localization patterns. Figure 16A illustrates the production of recombinant Dnmt3a proteins. His_6-tagged Dnmt3a, Dnmt3a:PC→AD, and Dnmt3a2 were expressed in E. coli and purified by metal chelation chromatography. The purity of the recombinant proteins was estimated by Coomassie blue staining (lanes 1-3) and their identity was verified by immunoblotting with antibody 64B1446 (lanes 4-6). Figure 16B illustrates methylation of double-stranded poly (dI-dC) by Dnmt3a and Dnmt3a2. The recombinant proteins were incubated with poly (dI-dC) in the presence of S-adenosyl-L-methionine [methyl-3H] and the methyltransferase activity was measured by the incorporation of ^{3}H-methyl group into poly (dI-dC). Each bar represents the mean + standard deviation of data from three independent reactions. Figure 16C demonstrates the localization of Dnmt3a and Dnmt3a2. GFP-Dnmt3a and Dnmt3a2 were transfected in NIH3T3 cells and the cells were fixed and analyzed by fluorescence microscopy. The top panel shows the GFP signal and the bottom panel shows the nuclei stained with DAPI. The arrows point to two heterochromatin regions and are used for orientation. Figure 16D illustrates the subcellular distribution of endogenous Dnmt3 proteins. ES cells were extracted to obtain the cytoplasmic, chromatin, and the nuclear matrix fractions (left). Equal amounts of each fraction were analyzed by immunoblotting with antibody 64B1446 (right, 1st panel), anti-histone H1 (2nd panel), and anti-lamin B (3rd panel).

[0031] Figures 17A-17D present Dnmt3a and Dnmt3b expression in embryoid bodies and mouse tissues. In Figure 17A undifferentiated ES cells (day 0) or differentiated embryoid bodies (day 2-14) were lysed and equal amount of proteins (30 μg/lane for Dnmt3a and tubulin, 5 μg/lane for Dnmt3a2 and Dnmt3b) were analyzed by immunoblotting with the indicated antibodies. In Figure 17B different organs from wild type or Dnmt3a$^{-/-}$ mice (3 weeks old) were homogenized and lysed, and the lysates immunoprecipitated and immunoblotted with Dnmt3a (64B1446) antibody (top panel) or Dnmt3b antibody 157 (bottom panel). ES cells were used as a positive control. Note that 64B1446 cross-reacts with a nonspecific band of ~105 kDa (indicated by *) in some tissues. Br, brain; Li, liver; Mu, muscle; Te, testis; Ht, heart; Sp, spleen; Th, thymus; St, stomach; Si, small intestine. In Figure 17C total RNA isolated from different tissues was analyzed by RT-PCR using primers either specific to *Dnmt3a* (F4 and R1) or to *Dnmt3a2* (F5 and R1). Lu, lung; Ov, ovary. In Figure 17D the same RNA samples were analyzed by RT-PCR using *Dnmt3*b-specific primers flanking exon 10 (top panel) or exons 21-22 (bottom panel) followed by Southern hybridization using Dnmt3b cDNA fragments as probes. *Dnmt3b1* and *Dnmt3b3* cDNAs were used as controls (lanes 1 and 2). The bands representing the presence (+) or absence (-) of exon 10 or exons 21-22 are indicated on the right and the major Dnmt3b isoforms present in ES cells and each tissue are indicated at the bottom.

[0032] Figures 18A-18D demonstrate that expression of DNMT3A2 and DNMT3B in human cell lines correlate with *de novo* methylation activity. Figures 18A-18B present expression of DNMT3A and DNMT3B in human EC cell lines. The indicated EC cell lines were lysed and equal amount of proteins (30 μg/lane) was analyzed by immunoblotting with antibody 64B1446 (A) or antibody 157 (B). Human DNMT3A and DNMT3B isoforms expressed in Cos-7 cells were used as positive controls. Figure 18C presents expression of DNMT1, DNMT3A, and DNMT3B in breast and ovarian tumor cell lines. For comparison, a human EC cell line, NCCIT, and mouse ES cells (J1) and NIH 3T3 cells were included (lanes 1, 11, 12). Equal amount of proteins (30 μg/lane) from the

indicated cell lysates was analyzed by immunoblotting with the indicated antibodies. Note that the anti-DNMT1 antibody does not recognize mouse Dnmt1. Figure 18D presents *De novo* methylation activity in human cell lines. The indicated cells were infected with Moloney Murine Leukemia Virus (MMLV). Five or 20 days after infection, genomic DNA was digested with *Kpn* I alone (K), *Kpn* I plus *Msp* I (K/M), or *Kpn* I plus *Hpa* II (K/H), and analyzed by Southern hybridization using the pMu3 probe. The MMLV and an enlarged 3' LTR region, two *Kpn* I (K) and five *Hpa* II/*Msp* I sites (vertical lines) and the pMu3 probe are shown at the bottom.

[0033] Figures 19A-19C demonstrate inactivation of Dnmt3a and Dnmt3b results in progressive loss of DNA methylation in ES cells.

(A) Genomic DNA from [*Dnmt3a-/-, Dnmt3b-/-*] ES cells (7aabb and 10aabb) that had been grown in culture for 5-40 passages, as well as wild-type (J1) and *Dnmt1* mutant (n/n and c/c) ES cells, was digested with HpaII and hybridized to probes for endogenous C-type retrovirus repeats (pMO), minor satellite repeats, and IAP repeats. As a control for complete digestion, DNA from J1 cells was digested with Msp I. The $Dnmt1^n$ allele (n stands for N-terminal disruption) is a partial loss-of-function mutation (Li, E., *et al.*, *Cell* 69:915-26 (1992)). and the $Dnmt1^c$ allele (c stands for disruption of the catalytic or C-terminal domain) is a null mutation (Lei, H., *et al.*, *Development* 122:3195-205 (1996)). (B) Genomic DNA from J1, *Dnmt3a-/-* (6aa), or *Dnmt3b-/-* (8bb) ES cells that had been grown in culture for 5-25 passages, as well as 7aabb (P40), was digested with HpaII and hybridized to pMO probe. (C) Lysates from the indicated ES cell lines were immunoblotted with anti-Dnmt1 and anti-tubulin antibodies.

[0034] Figures 20A-20B present stable expression of Dnmt3a and Dnmt3b isoforms in late-passage 7aabb cells. (A) Schematic diagram of Dnmt3a and Dnmt3b isoforms. The conserved PWWP and PHD domains, the methyltransferase motifs (I, IV, VI, IX, and X), and the sites of alternative splicing are indicated (the C-terminal 45 amino acids of Dnmt3b5 are out of frame and shown as an open bar). The locations of the epitopes for the Dnmt3a

and Dnmt3b antibodies are also shown. (B) cDNAs encoding Dnmt3a/3b isoforms were subcloned in an expression vector (schematically shown at the top) and these constructs were individually electroporated into late-passage (P70) 7aabb cells, which were subsequently selected in blasticidin-containing medium for seven days. Blasticidin-resistant clones were analyzed with immunoblotting using anti-Dnmt3a (middle panel) or anti-Dnmt3b (bottom panel) antibodies. As a loading control, the same membranes were immunoblotted with anti-tubulin antibody.

[0035] Figures 21A-21I demonstrate that expression of Dnmt3a/3b proteins in 7aabb cells restores DNA methylation. (A-D) Methylation of repetitive sequences. Genomic DNA from the indicated ES cell lines was digested with Hpa II (A-C) or Mae II (D) and hybridized to the indicated probes. DNA from J1 cells digested with Msp I was used as a control for complete digestion. (E) Analysis of the methylation status of the major satellite repeating unit by bisulfite sequencing. Genomic DNA from J1 and 7aabb cells as well as stable cell lines expressing Dnmt3a, Dnmt3a2, Dnmt3b1, and Dnmt3b3 was analyzed. The methylation status of six CpG sites from 8-12 individual clones is shown schematically (black circles represent methylated sites), and the percentages of methylated CpG sites are indicated in parenthesis. (F-I) Methylation of unique genes. The same genomic DNA samples described in (A-D) were digested with Bam HI and Hha I (F and H), EcoRI and Hpa II (G), or EcoRV and Hha I (I) and hybridized to probes corresponding to the 3' region of *β-globin* (F), the 5' region of *Pgk-1* (G), an exon of *Pgk-2* (H), or the 5' region of *Xist* (I). DNA from J1 cells digested with Bam HI alone (F and H) or EcoRI alone (G) was used as controls.

[0036] Figures 22A-22E demonstrate expression of Dnmt3a and Dnmt3b proteins in 7aabb cells fails to restore maternal methylation imprints. The same DNA samples described in Fig. 3 were digested with Sac I and Hha I (A), Bam HI and Hpa II (B), Pvu II and Hpa II (C and D), or Xba I and Hha I (E) and hybridized to probes corresponding to the 5' upstream region of *H19* (A), the DMR2 of *Igf2*

(B), region 2 of *Igf2r* (C), the DMR of *Peg1*, or the DMR1 of *Snrpn* (E). As controls, DNA from J1 cells was digested with the corresponding enzymes without Hha I or Hpa II. The fragments derived from the paternal (p) and maternal (m) alleles are indicated.

[0037] Figures 23A-23E demonstrate Dnmt3b6 has no enzymatic activity in vivo. (A) Strategy of targeted deletion of Dnmt3b exons 21 and 22. The top line shows the Dnmt3b genomic structure with exons represented by vertical bars. The targeting vector (second line) was constructed by replacing exons 21 and 22 with a PGK-puromycin cassette. A PGK-DTA cassette was introduced for negative selection to increase the targeting frequency. (B) Southern analysis of the genotype of ES cell lines. Genomic DNA was digested with EcoRV and hybridized to a 3' external probe, as shown in (A). The 16-kb wild-type allele, the 5-kb Dnmt3b1 targeted allele, and the 14-kb Dnmt3b null allele (30) are indicated. (C) Lysates from the indicated cell lines were immunoblotted with anti-Dnmt3b (top), anti-Dnmt3a (middle), and anti-tubulin (bottom) antibodies. (D and E) Genomic DNA from the indicated ES cell lines was digested with Hpa II and hybridized to probes for endogenous C-type retrovirus repeats (D) and minor satellite repeats (E).

[0038] Figures 24A-24B demonstrate Dnmt3b3 inhibits de novo methylation by Dnmt3a and Dnmt3b. (A) Dnmt3a, Dnmt3a2, or Dnmt3b1 cDNA was electroporated into late-passage 7aabb cells in the presence or absence of Dnmt3b3 cDNA, and stable clones were analyzed for protein expression by immunoblotting using anti-Dnmt3a (top), anti-Dnmt3b (middle), and anti-tubulin (bottom) antibodies. (B) Genomic DNA from the indicated stable clones was analyzed for methylation using pMO, Igf2, and Xist probes, as indicated.

[0039] Figures 25A-25B demonstrate active Dnmt3a/3b isoforms rescue the capacity of late-passage 7aabb cells to form terotomas in nude mice. (A) The indicated ES cell lines were injected into nude mice subcutaneously on both sides (3-4 mice for each cell line, $5x10^5$ cells per site) and the mice were examined for terotomas after 4 weeks. A typical representation of the size of the terotomas

derived from each cell line is shown. (B) Histological sections of teratomas derived from J1, early-passage (P10) 7aabb, and Dnmt3a, Dnmt3a2, and Dnmt3b1 stable clones showing the presence of multiple types of differentiated cells.

[0040] Figures 26A-26C demonstrate Dnmt1 and Dnmt3 proteins function cooperatively in maintaining methylation patterns. (A) Dnmt1 or Dnmt3a was overexpressed in 7aabb (P70) or Dnmt1-/- (c/c) ES cells as indicated and stable clones were examined for protein expression by immunoblotting using anti-Dnmt1 (top), anti-Dnmt3a (middle), and anti-tubulin (bottom) antibodies. (B and C) Genomic DNA from the indicated ES cell lines was analyzed for methylation of repetitive sequences (B) and unique genes (C) using the indicated probes.

[0041] Figure 27 presents mouse Dnmt3a2 promoter sequence. Underlined sequences represent GC-rich regions that have high promoter potential as predicted by the computer program PROSCAN. An about 100 to 250 nucleotide region is represented by 250 "N" nucleotides from nucleotide position 723-972. This region could not be sequenced, presumably due to high GC content. The sequence of the first exon of Dnmt3a2 is italicized and bolded.

[42] Figure 28 presents human DNMT3A2 promoter sequence. The sequence of the first exon of DNMT3A2 is italicized. The promoter sequence was identified by BLAST searching SEQ ID NO:118 against the human genome sequence database available at http://www.ncbi.nlm.nih.gov/BLAST/. The sequence of the first exon of DNMT3A2 is italicized and bolded.

[43] Figure 29 presents a sequence alignment of mouse Dnmt3a2 and human DNMT3A2 promoter sequence. The about 100 to about 250 nucleotide region in the mouse Dnmt3a2 promoter, denoted by 250 "N" nucleotides in Figure 27, was not counted in the numbering of the nucleotides.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Definitions

[0044] In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

[0045] ***Cloning vector:*** A plasmid or phage DNA or other DNA sequence which is able to replicate autonomously in a host cell, and which is characterized by one or a small number of restriction endonuclease recognition sites at which such DNA sequences may be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a DNA fragment may be spliced in order to bring about its replication and cloning. The cloning vector may further contain a marker suitable for use in the identification of cells transformed with the cloning vector. Markers, for example, provide tetracycline resistance or ampicillin resistance.

[0046] ***Expression vector***: A vector similar to a cloning vector but which is capable of enhancing the expression of a gene which has been cloned into it, after transformation into a host. The cloned gene is usually placed under the control of (i.e., operably linked to) certain control sequences such as promoter sequences. Promoter sequences may be either constitutive or inducible.

[0047] ***Recombinant Host:*** According to the invention, a recombinant host may be any prokaryotic or eukaryotic host cell which contains the desired cloned genes on an expression vector or cloning vector. This term is also meant to include those prokaryotic or eukaryotic cells that have been genetically engineered to contain the desired gene(s) in the chromosome or genome of that organism. For examples of such hosts, *see* Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989). Preferred recombinant hosts are eukaryotic cells transformed

with the DNA construct of the invention. More specifically, mammalian cells are preferred.

[0048] ***Recombinant vector:*** Any cloning vector or expression vector which contains the desired cloned gene(s).

[0049] ***Host Animal:*** Transgenic animals, all of whose germ and somatic cells contain the DNA construct of the invention. Such transgenic animals are in general vertebrates. Preferred host animals are mammals such as non-human primates, humans, mice, sheep, pigs, cattle, goats, guinea pigs, rodents, e.g. rats, and the like. The term host animal also includes animals in all stages of development, including embryonic and fetal stages.

[0050] ***Promoter:*** A DNA sequence generally described as the 5′ region of a gene, located proximal to the start codon. The transcription of an adjacent gene(s) is initiated at the promoter region. If a promoter is an inducible promoter, then the rate of transcription increases in response to an inducing agent. In contrast, the rate of transcription is not regulated by an inducing agent if the promoter is a constitutive promoter. According to the invention, preferred promoters are heterologous to the *de novo* DNA cytosine methyltransferase genes, that is, the promoters do not drive expression of the gene in a mouse or human. Such promoters include the CMV promoter (InVitrogen, San Diego, CA), the SV40, MMTV, and hMTIIa promoters (U.S. 5,457,034), the HSV-1 4/5 promoter (U.S. 5,501,979), and the early intermediate HCMV promoter (WO92/17581). In one emdodiment, it is preferred that the promoter is tissue-specific, that is, it is induced selectively in a specific tissue. Also, tissue-specific enhancer elements may be employed. Additionally, such promoters may include tissue and cell-specific promoters of an organism.

[0051] ***Gene:*** A DNA sequence that contains information needed for expressing a polypeptide or protein.

[0052] ***Structural gene:*** A DNA sequence that is transcribed into messenger RNA (mRNA) that is then translated into a sequence of amino acids characteristic of a specific polypeptide.

[0053] ***Complementary DNA (cDNA):*** A "complementary DNA," or "cDNA" gene includes recombinant genes synthesized by reverse transcription of mRNA and from which intervening sequences (introns) have been removed.

[0054] ***Expression:*** Expression is the process by which a polypeptide is produced from a structural gene. The process involves transcription of the gene into mRNA and the translation of such mRNA into polypeptide(s).

[0055] ***Homologous/Nonhomologous:*** Two nucleic acid molecules are considered to be "homologous" if their nucleotide sequences share a similarity of greater than 40%, as determined by HASH-coding algorithms (Wilber, W.J. and Lipman, D.J., *Proc. Natl. Acad. Sci. 80*:726-730 (1983)). Two nucleic acid molecules are considered to be "nonhomologous" if their nucleotide sequences share a similarity of less than 40%.

[0056] ***Polynucleotide:*** This term generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications have been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

[0057] ***Isoform***: This term refers to a protein or polynucleotide that is produced from an alternatively spliced RNA transcript or from an RNA transcript that is generated by an alternative promoter. As used herein, "isoform" refers to the polypeptides and polynucleotides encoding the polypeptides.

[0058] ***Polypeptide:*** This term refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation,

racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. See, for instance, *Proteins-Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.,,* "Analysis for protein modifications and nonprotein cofactors", *Methods in Enzymol. 182*:626-646 (1990) and Rattan *et al.,,* "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci 663*:48-62 (1992).

[0059] ***Variant***: The term used herein is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

[0060] ***Identity:*** This term refers to a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that

the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. (See, *e.g.*: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al.,, Nucleic Acids Research 12(I)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.,, J Mol. Biol 215*:403 (1990)).

[0061] Therefore, as used herein, the term "identity" represents a comparison between a test and reference polynucleotide. More specifically, reference polynucleotides are identified in this invention as SEQ ID NOS: 1, 2, 3, 4, 83, and 84 and a test polynucleotide is defined as any polynucleotide that is 90% or more identical to a reference polynucleotide. As used herein, the term "90% or more" refers to percent identities from 90 to 99.99 relative to the reference polynucleotide. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 nucleotides, that no more than 10% (*i.e.*, 10 out of 100) nucleotides in the test polynucleotide differ from that of the reference polynucleotide. Such

differences may be represented as point mutations randomly distributed over the entire length of the sequence or they may be clustered in one or more locations of varying length up to the maximum allowable 10 nucleotide difference. Differences are defined as nucleotide substitutions, deletions or additions of sequence. These differences may be located at any position in the sequence, including but not limited to the 5' end, 3' end, coding and non coding sequences.

[0062] ***Fragment:*** A "fragment" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to any polypeptide subset of that molecule.

[0063] ***Functional Derivative:*** The term "functional derivatives" is intended to include the "variants," "analogues," or "chemical derivatives" of the molecule. A "variant" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to a naturally occurring molecule substantially similar to either the entire molecule, or a fragment thereof. An "analogue" of a molecule such as *de novo* DNA cytosine methyltransferases is meant to refer to a non-natural molecule substantially similar to either the entire molecule or a fragment thereof.

[0064] A molecule is said to be "substantially similar" to another molecule if the sequence of amino acids in both molecules is substantially the same, and if both molecules possess a similar biological activity. Thus, provided that two molecules possess a similar activity, they are considered variants as that term is used herein even if one of the molecules contains additional amino acid residues not found in the other, or if the sequence of amino acid residues is not identical.

[0065] As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half-life, etc. The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, etc. Examples of moieties capable of mediating such effects are disclosed in *Remington's Pharmaceutical Sciences* (1980) and will be apparent to those of ordinary skill in the art.

[0066] ***Protein Activity* or *Biological Activity of the Protein:*** These expressions refer to the metabolic or physiologic function of *de novo* DNA cytosine methyltransferase protein including similar activities or improved activities or these activities with decreased undesirable side-effects. Also included are antigenic and immunogenic activities of said *de novo* DNA cytosine methyltransferase protein. Among the physiological or metabolic activities of said protein is the transfer of a methyl group to the cytosine C5 position of duplex DNA. Such DNA may completely lack any methylation of may be hemimethylated. As demonstrated in Examples 4 and 5, *de novo* DNA cytosine methyltransferases methylate C5 in cytosine moieties in nonmethylated DNA.

[0067] ***De novo DNA Cytosine Methyltransferase Polynucleotides:*** This term refers to a polynucleotide containing a nucleotide sequence that encodes a *de novo* DNA cytosine methyltransferase polypeptide or fragment thereof, variant, or isoform or that encodes a *de novo* DNA cytosine methyltransferase polypeptide or fragment thereof, variant, or isoform, wherein said nucleotide sequence has at least 90% identity to a nucleotide sequence encoding the polypeptide of SEQ ID Nos: 5, 6, 7, 8, 85 or 86 or a corresponding fragment thereof, or which has sufficient identity to a nucleotide sequence contained in SEQ ID NO:1, 2, 3, 4, 83, or 84.

[0068] *De novo DNA Cytosine Methyltransferase Polypeptides:* This term refers to polypeptides with amino acid sequences sufficiently similar to the *de novo* DNA cytosine methyltransferase protein sequence in SEQ ID NO:5, 6, 7, 8, 85 or 86 and that at least one biological activity of the protein is exhibited.

[0069] ***Antibodies:*** As used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

[0070] ***Substantially pure:*** As used herein means that the desired purified protein is essentially free from contaminating cellular components, said components being associated with the desired protein in nature, as evidenced by

a single band following polyacrylamide-sodium dodecyl sulfate gel electrophoresis. Contaminating cellular components may include, but are not limited to, proteinaceous, carbohydrate, or lipid impurities.

[0071] The term "substantially pure" is further meant to describe a molecule which is homogeneous by one or more purity or homogeneity characteristics used by those of skill in the art. For example, a substantially pure *de novo* DNA cytosine methyltransferases will show constant and reproducible characteristics within standard experimental deviations for parameters such as the following: molecular weight, chromatographic migration, amino acid composition, amino acid sequence, blocked or unblocked N-terminus, HPLC elution profile, biological activity, and other such parameters. The term, however, is not meant to exclude artificial or synthetic mixtures of the factor with other compounds. In addition, the term is not meant to exclude *de novo* DNA cytosine methyltransferase fusion proteins isolated from a recombinant host.

[0072] ***Isolated:*** A term meaning altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of a *de novo* DNA cytosine methyltransferase polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67*:31-40 (1988).

[0073] ***Neoplastic disorder:*** This term refers to a disease state which is related to the hyperproliferation of cells. Neoplastic disorders include, but are not limited to, carcinomas, sarcomas and leukemia.

[0074] ***Gene Therapy:*** A means of therapy directed to altering the normal pattern of gene expression of an organism. Generally, a recombinant polynucleotide is introduced into cells or tissues of the organism to effect a change in gene expression.

[0075] ***Antisense RNA gene/Antisense RNA.*** In eukaryotes, mRNA is transcribed by RNA polymerase II. However, it is also known that one may construct a gene containing a RNA polymerase II template wherein a RNA sequence is transcribed which has a sequence complementary to that of a specific mRNA but is not normally translated. Such a gene construct is herein termed an "antisense RNA gene" and such a RNA transcript is termed an "antisense RNA." Antisense RNAs are not normally translatable due to the presence of translation stop codons in the antisense RNA sequence.

[0076] ***Antisense oligonucleotide:*** A DNA or RNA molecule or a derivative of a DNA or RNA molecule containing a nucleotide sequence which is complementary to that of a specific mRNA. An antisense oligonucleotide binds to the complementary sequence in a specific mRNA and inhibits translation of the mRNA. There are many known derivatives of such DNA and RNA molecules. See, for example, U.S. Patent Nos. 5,602,240, 5,596,091, 5,506,212, 5,521,302, 5,541,307, 5,510,476, 5,514,787, 5,543,507, 5,512,438, 5,510,239, 5,514,577, 5,519,134, 5,554,746, 5,276,019, 5,286,717, 5,264,423, as well as WO96/35706, WO96/32474, WO96/29337 (thiono triester modified antisense oligodeoxynucleotide phosphorothioates), WO94/17093 (oligonucleotide alkylphosphonates and alkylphosphothioates), WO94/08004 (oligonucleotide phosphothioates, methyl phosphates, phosphoramidates, dithioates, bridged phosphorothioates, bridge phosphoramidates, sulfones, sulfates, ketos, phosphate esters and phosphorobutylamines (van der Krol *et al.,, Biotech. 6*:958-976 (1988); Uhlmann *et al.,, Chem. Rev. 90*:542-585 (1990)), WO94/02499 (oligonucleotide

alkylphosphonothioates and arylphosphonothioates), and WO92/20697 (3'-end capped oligonucleotides). Particular *de novo* DNA cytosine methyltransferase antisense oligonucleotides of the present invention include derivatives such as S-oligonucleotides (phosphorothioate derivatives or S-oligos, *see*, Jack Cohen, *Oligodeoxynucleotides, Antisense Inhibitors of Gene Expression*, CRC Press (1989)). S-oligos (nucleoside phosphorothioates) are isoelectronic analogs of an oligonucleotide (O-oligo) in which a nonbridging oxygen atom of the phosphate group is replaced by a sulfur atom. The S-oligos of the present invention may be prepared by treatment of the corresponding O-oligos with 3*H*-1,2-benzodithiol-3-one-1,1-dioxide which is a sulfur transfer reagent. *See* Iyer *et al.,, J. Org. Chem. 55*:4693-4698 (1990); and Iyer *et al.,, J. Am. Chem. Soc. 112*:1253-1254 (1990).

[0077] ***Antisense Therapy:*** A method of treatment wherein antisense oligonucleotides are administered to a patient in order to inhibit the expression of the corresponding protein.

I. Deposited Material

[0078] The invention relates to polynucleotides encoding and polypeptides of novel *de novo* DNA cytosine methyltransferase proteins. The invention relates especially to *de novo* DNA cytosine methyltransferase mouse Dnmt3a, Dnmt3a2 and Dnmt3b cDNAs and the human DNMT3A, DNMT3A2 and DNMT3B cDNAs set out in SEQ ID NOs:1, 83, 2, 3, 84 and 4, respectively. The invention also relates to mouse Dnmt3a, Dnmt3a2 and Dnmt*3*b and human DNMT3A, DNMT3A2 and DNMT3B *de novo* DNA cytosine methyltransferase polypeptides set out in SEQ ID NOs:5, 85, 6, 7, 86 and 8, respectively. The invention further relates to the *de novo* DNA cytosine methyltransferase nucleotide sequences of the mouse Dnmt3a cDNA (plasmid pMT3a), Dnmt3a2 cDNA, and Dnmt3b cDNA (plasmid pMT3b), and the human *DNMT*3A cDNA (plasmid pMT3A), and DNMT3A2 cDNA in ATCC Deposit Nos.209933, PTA-4611, 209934, 98809, and PTA-4610 respectively, and the amino acid sequences encoded

therein. The invention further relates to *de novo* DNA cytosine methyltransferase promoter sequences of mouse Dnmt3a2 (plasmid P2-luc), and human DNMT3A2 (plasmid P2-luc-human) deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia 20110-2209, USA, on _____, and assigned ATCC Deposit Nos. _____ and _____.

[0079] The nucleotide sequence of the human *DNMT3B* cDNA identified in SEQ ID NO:4 is available in a clone (ATCC Deposit No. 326637) independently deposited by the I.M.A.G.E. Consortium. The invention relates to the *de novo* DNA cytosine methyltransferase polypeptide encoded therein.

[0080] Clones containing mouse Dnmt3a and Dnmt3b cDNAs were deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia 20110-2209, USA, on June 16, 1998, and assigned ATCC Deposit Nos. 209933 and 209934, respectively. The human DNMT3A cDNA was deposited with the ATCC on July 10, 1998, and assigned ATCC Deposit No. 98809. Clones containing mouse Dnmt3a2 and human DNMT3A2 were deposited with the American Type Culture Collection (ATCC) on August 23, 2002 and assigned ATCC deposit No. PTA-4611 and PTA-4610, respectively.

[0081] While the ATCC deposits are believed to contain the *de novo* DNA cytosine methyltransferase cDNA sequences shown in SEQ ID NOs:1, 2, 3, 4, 83 and 84, the nucleotide sequences of the polynucleotide contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

[0082] The deposits for mouse Dnmt3a, Dnmt3a2 and Dnmt3b cDNAs and the human DNMT3A and DNMT3A2 cDNA were made under the terms of the Budapest Treaty on the international recognition of the deposit of micro-organisms for purposes of patent procedure. The deposits are provided merely as a convenience for those of skill in the art and are not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. § 112.

II. Polynucleotides of the Invention

[0083] Another aspect of the invention relates to isolated polynucleotides, and polynucleotides closely related thereto, which encode the *de novo* DNA cytosine methyltransferase polypeptides. As shown by the results presented in Figure 5, sequencing of the cDNAs contained in the deposited clones encoding mouse and human *de novo* DNA cytosine methyltransferases confirms that the *de novo* DNA cytosine methyltransferase proteins of the invention are structurally related to other proteins of the DNA methyltransferase family.

[0084] The polynucleotides of the present invention encoding *de novo* DNA cytosine methyltransferase proteins may be obtained using standard cloning and screening procedures as described in Examples 1 and 5. Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

[0085] Among particularly preferred embodiments of the invention are polynucleotides encoding *de novo* DNA cytosine methyltransferase polypeptides having the amino acid sequence set out in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85, or SEQ ID NO:86, and variants thereof.

[0086] A particular nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide may be identical over its entire length to the coding sequence in SEQ ID NOs:1, 2, 3, 83, or 84. Alternatively, a particular nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide may be an alternate form of SEQ ID NOs:1, 2, 3, 4, 83, or 84 due to degeneracy in the genetic code or variation in codon usage encoding the polypeptides of SEQ ID NOs:5, 6, 7, 8, 85, or 86. Preferably, the polynucleotides of the invention contain a nucleotide sequence that is highly identical, at least 90% identical, with a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide or at least 90% identical with the encoding nucleotide sequence set

forth in SEQ ID NOs:1, 2, 3, 83, or 84. Polynucleotides of the invention may be 90 to 99% identical to the nucleotides sequence set forth in SEQ ID NO:4.

[0087] When a polynucleotide of the invention is used for the recombinant production of a *de novo* DNA cytosine methyltransferase polypeptide, the polynucleotide may include the coding sequence for the full-length polypeptide or a fragment thereof, by itself; the coding sequence for the full-length polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.,*, *Proc Natl Acad Sci USA 86*:821-824 (1989), or it may be the HA tag, which corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al.,*, *Cell 37*:767, 1984). The polynucleotide may also contain non-coding 5′ and 3′ sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

[0088] Embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 90% identical, and more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3; SEQ ID NO:83, or SEQ ID NO:84; (b) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809, ATCC Deposit No. PTA-4611, or ATCC Deposit No. PTA-4610; or (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b). Additionally, an isolated nucleic acid of the invention

may be a polynucleotide at least 90% but not more than 99% identical to (a) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence in SEQ ID NO:4; (b) a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No.326637; or (c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

[0089] Conventional means utilizing known computer programs such as the BestFit program (Wisconsin Sequence Analysis Package, Version 10 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) may be utilized to determine if a particular nucleic acid molecule is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identical to any one of the nucleotide sequences shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:83, or SEQ ID NO:84 or to any one of the nucleotide sequences of the deposited cDNA clones contained in ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809, ATCC Deposit No. 326637, ATCC Deposit No. PTA-4611, or ATCC Deposit No. PTA-4610, respectively.

[0090] Further preferred embodiments are polynucleotides encoding *de novo* DNA cytosine methyltransferases and *de novo* DNA cytosine methyltransferase variants that have an amino acid sequence of the *de novo* DNA cytosine methyltransferase protein of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85, or SEQ ID NO:86 in which several, 1, 1-2, 1-3, 1-5 or 5-10 amino acid residues are substituted, deleted or added, in any combination.

[0091] Further preferred embodiments of the invention are polynucleotides that are at least 90% identical over their entire length to a polynucleotide encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence set out in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85, or SEQ ID NO:86, and polynucleotides which are complementary to such polynucleotides. Most highly preferred are

polynucleotides that comprise regions that are at least 90% identical over their entire length to a polynucleotide encoding the *de novo* DNA cytosine methyltransferase polypeptides of the ATCC deposited human DNMT3A and DNMT3A2 cDNA clones and polynucleotides complementary thereto, and 90% to 99% identical over their entire length to a polynucleotide encoding the *de novo* DNA cytosine methyltransferase polypeptides of the ATCC deposited human DNMT3B cDNA clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 95% identical over their entire length to the same are particularly preferred, and those with at least 97% identity are especially preferred. Furthermore, those with at least 98% identity are highly preferred and with at least 99% identity being the most preferred.

[0092] In a more specific embodiment, the nucleic acid molecules of the present invention, *e.g.*, isolated nucleic acids comprising a polynucleotide having a nucleotide sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide or fragment thereof, are not the sequence of nucleotides, the nucleic acid molecules (*e.g.*, clones), or the nucleic acid inserts identified in one or more of the below cited public EST or STS GenBank Accession Reports.

[0093] The following public ESTs were identified that relate to portions of SEQ ID NO:1: AA052791(SEQ ID NO:9); AA111043(SEQ ID NO:10); AA154890(SEQ ID NO:11); AA240794(SEQ ID NO:12); AA756653(SEQ ID NO:13); W58898(SEQ ID NO:14); W59299(SEQ ID NO:15); W91664(SEQ ID NO:16); W91665(SEQ ID NO:17); to portions of SEQ ID NO:2: AA116694 (SEQ ID NO:18); AA119979 (SEQ ID NO:19); AA177277 (SEQ ID NO:20); AA210568 (SEQ ID NO:21); AA399749 (SEQ ID NO:22); AA407106 (SEQ ID NO:23); AA575617 (SEQ ID NO:24); to portions of SEQ ID NO:3: AA004310 (SEQ ID NO:25); AA004399 (SEQ ID NO:26); AA312013 (SEQ ID NO:27); AA355824 (SEQ ID NO:28); AA533619 (SEQ ID NO:29); AA361360 (SEQ ID NO:30); AA364876 (SEQ ID NO:31); AA503090 (SEQ ID NO:32); AA533619 (SEQ ID NO:33); AA706672 (SEQ ID NO:34); AA774277 (SEQ ID NO:35); AA780277 (SEQ ID NO:36); H03349 (SEQ ID NO:37); H04031 (SEQ ID

NO:38); H53133 (SEQ ID NO:39); H53239 (SEQ ID NO:40); H64669 (SEQ ID NO:41); N26002 (SEQ ID NO:42); N52936 (SEQ ID NO:43); N88352 (SEQ ID NO:44); N89594 (SEQ ID NO:45); R19795 (SEQ ID NO:46); R47511 (SEQ ID NO:47); T50235 (SEQ ID NO:48); T78023 (SEQ ID NO:49); T78186 (SEQ ID NO:50); W22886 (SEQ ID NO:51); W67657 (SEQ ID NO:52); W68094 (SEQ ID NO:53); W76111 (SEQ ID NO:54); Z38299 (SEQ ID NO:55); Z42012 (SEQ ID NO:56); and that relate to SEQ ID NO:4: AA206103(SEQ ID NO:57); AA206264(SEQ ID NO:58); AA216527(SEQ ID NO:59); AA216697(SEQ ID NO:60); AA305044(SEQ ID NO:61); AA477705(SEQ ID NO:62); AA477706(SEQ ID NO:63); AA565566(SEQ ID NO:64); AA599893(SEQ ID NO:65); AA729418(SEQ ID NO:66); AA887508(SEQ ID NO:67); F09856(SEQ ID NO:68); F12227(SEQ ID NO:69); N39452(SEQ ID NO:70); N48564(SEQ ID NO:71); T66304(SEQ ID NO:72); and T66356(SEQ ID NO:73); AA736582(SEQ ID NO:77); AA748883(SEQ ID NO:78); AA923295(SEQ ID NO:79); AAI000396(SEQ ID NO:80); AI332472(SEQ ID NO:81); W22473(SEQ ID NO:82) and the I.M.A.G.E. Consortium clone ID 22089 (ATCC Deposit No. 326637)(SEQ ID NO:76). Additionally, STSs G06200(SEQ ID NO:74) and G15302(SEQ ID NO:75) were identified in a search with SEQ ID NOS.:3 and 4, respectively. All identified public sequences are hereby incorporated by reference.

[0094] Polynucleotides of the invention also include isoforms of the mouse Dnmt3a and human DNMT3A sequences disclosed herein which may arise through the use of an alternative promoter of the Dnmt3a or DNMT3A gene. For example, isoforms of mouse Dnmt3a arising through differential promoter usage include but are not limited to a polynucleotide represented by SEQ ID NO:83. Isoforms of human DNMT3A arising through differential promoter usage include but are not limited to the polynuclotide represented by SEQ ID NO:84.

[0095] The present invention is further directed to fragments of SEQ ID NO:1, 2, 3, 83 or 84, or to fragments of the cDNA nucleotide sequence found in ATCC Deposit Nos. 209933, 209934, 98809, PTA-4611, or PTA-4610. A fragment may

be defined to be at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. Such fragments are useful as diagnostic probes and primers as discussed herein. Of course larger DNA fragments are also useful according to the present invention, as are fragments corresponding to most, if not all, of the nucleotide sequence of the cDNA clones contained in the plasmids deposited as ATCC Deposit No. 209933, ATCC Deposit No. 209934 ATCC Deposit No. 98809, ATCC Deposit No. PTA-4611, ATCC Deposit No. PTA-4610 or as shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:83, or SEQ ID NO:84. Generally, polynucleotide fragments of the invention may be defined algebraically in the following way: (a) for SEQ ID NO:1, as 15 + N, wherein N equals zero or any positive integer up to 4176; (b) for SEQ ID NO:2, as 15 + N, wherein N equals zero or any positive integer up to 4180; and (c) for SEQ ID NO:3, as 15 + N, wherein N equals zero or any positive integer up to 4401; (d) for SEQ ID NO:83, as 15 + N, wherein N equals zero or any positive integer up to 2303; (e) for SEQ ID NO:84, as 15 + N, wherein N equals zero or any positive integer up to 2356. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from a nucleotide sequence of the ATCC deposited cDNAs or the nucleotide sequence as shown in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:83 or SEQ ID NO:84.

[0096] In a specific embodiment, the fragments of SEQ ID NO:1 and SEQ ID NO:2 are SEQ ID NO:83 and SEQ ID NO:84, respectively.

[0097] In another embodiment, the invention is directed to fragments of SEQ ID NO:4. Such fragments are defined as comprising the nucleotide sequence encoding the specific amino acid residues integral and immediately adjacent to the site where DNMT3B exons are spliced together. The DNMT3B sequence of SEQ ID NO:4 consists of 23 exon sequences defined accordingly: Exon 1 consists of nucleotides 1-108 of SEQ ID NO:4; Exon 2 consists of nucleotides 109-256 of SEQ ID NO:4; Exon 3 consists of nucleotides 257-318 of SEQ ID NO:4; Exon 4 consists of nucleotides 319-420 of SEQ ID NO:4; Exon 5 consists

of nucleotides 421-546 of SEQ ID NO:4; Exon 6 consists of nucleotides 547-768 of SEQ ID NO:4; Exon 7 consists of nucleotides 769-927 of SEQ ID NO:4; Exon 8 consists of nucleotides 928-1035 of SEQ ID NO:4; Exon 9 consists of nucleotides 1036-1180 of SEQ ID NO:4; Exon 10 consists of nucleotides 1181-1240 of SEQ ID NO:4; Exon 11 consists of nucleotides 1241-1366 of SEQ ID NO:4; Exon 12 consists of nucleotides 1367-1411 of SEQ ID NO:4; Exon 13 consists of nucleotide 1412-1491 of SEQ ID NO:4; Exon 14 consists of nucleotides 1492-1604 of SEQ ID NO:4; Exon 15 consists of nucleotides 1605-1788 of SEQ ID NO:4; Exon 16 consists of nucleotides 1789-1873 of SEQ ID NO:4; Exon 17 consists of nucleotides 1874-2019 of SEQ ID NO:4; Exon 18 consists of nucleotides 2020-2110 of SEQ ID NO:4; Exon 19 consists of nucleotides 2111-2259 of SEQ ID NO:4; Exon 20 consists of nucleotides 2260-2345 of SEQ ID NO:4; Exon 21 consists of nucleotides 2346-2415 of SEQ ID NO:4; Exon 22 consists of nucleotides 2416-2534 of SEQ ID NO:4; and Exon 23 consists of nucleotides 2535-4145 of SEQ ID NO:4.

[0098] It should be understood by those skilled in the art that with regards to SEQ ID NO:4, Exon 1 and Exon 23 are herein defined for the purposes of the invention. The first nucleotide of Exon 1 may or may not be the transcriptional start site for the DNMT3B genomic locus, and the last nucleotide identified for Exon 23 may or may not reflect the last nucleotide transcribed *in vivo*.

[0099] Thus, by way of example, fragments of SEQ ID NO:4 comprise the following exon-exon junctions of 20 nucleotides in length: the exon1/exon 2 junction of nucleotides 98-118 of SEQ ID NO:4; the exon 2/exon 3 junction of nucleotides 246-266 of SEQ ID NO:4; the exon 3/exon 4 junction of nucleotides 308-328 of SEQ ID NO:4; the exon 4/exon 5 junction of nucleotides 410-430 of SEQ ID NO:4; the exon 5/exon 6 junction of nucleotides 536-556 of SEQ ID NO:4; the exon 6/exon 7 junction of nucleotides 758-778 of SEQ ID NO:4; the exon 7/exon 8 junction of nucleotides 917-937 of SEQ ID NO:4; the exon 8/exon 9 junction of nucleotides 1025-1045 of SEQ ID NO:4; the exon 9/exon 10 junction of nucleotides 1170-1190 of SEQ ID NO:4; the exon 10/exon 11

junction of nucleotides 1230-1250 of SEQ ID NO:4; the exon 11/exon 12 junction of nucleotides 1356-1376 of SEQ ID NO:4; the exon 12/exon 13 junction of nucleotides 1401-1421 of SEQ ID NO:4; the exon 13/exon 14 junction of nucleotides 1481-1501 of SEQ ID NO:4; the exon 14/exon 15 junction of nucleotides 1594-1614 of SEQ ID NO:4; the exon 15/exon 16 junction of nucleotides 1778-1798 of SEQ ID NO:4; the exon 16/exon 17 junction of nucleotides 1863-1883 of SEQ ID NO:4; the exon 17/exon 18 junction of nucleotides 2009-2029 of SEQ ID NO:4; the exon 18/exon 19 junction of nucleotides 2100-2120 of SEQ ID NO:4; the exon 19/exon 20 junction of nucleotides 2249-2269 of SEQ ID NO:4; the exon 20/exon 21 junction of nucleotides 2335-2355 of SEQ ID NO:4; the exon 21/exon 22 junction of nucleotides 2405-2425 of SEQ ID NO:4; and the exon 22/exon 23 junction of nucleotides 2524-2544 of SEQ ID NO:4.

[0100] As will be clear to those skilled in the art, other exon-exon junction fragments of SEQ ID NO:4 are possible which comprise 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, etc., nucleotides of SEQ ID NO:4. For the purposes of constructing such fragments, the following exon-exon junctions are identified: the exon1/exon 2 junction of nucleotides 108 and 109 of SEQ ID NO:4; the exon 2/exon 3 junction of nucleotides 256 and 257 of SEQ ID NO:4; the exon 3/exon 4 junction of nucleotides 318 and 319 of SEQ ID NO:4; the exon 4/exon 5 junction of nucleotides 420 and 421 of SEQ ID NO:4; the exon 5/exon 6 junction of nucleotides 546 and 547 of SEQ ID NO:4; the exon 6/exon 7 junction of nucleotides 768 and 769 of SEQ ID NO:4; the exon 7/exon 8 junction of nucleotides 927 and 928 of SEQ ID NO:4; the exon 8/exon 9 junction of nucleotides 1035 and 1036 of SEQ ID NO:4; the exon 9/exon 10 junction of nucleotides 1180 and 1181 of SEQ ID NO:4; the exon 10/exon 11 junction of nucleotides 1240 and 1241 of SEQ ID NO:4; the exon 11/exon 12 junction of nucleotides 1366 and 1367 of SEQ ID NO:4; the exon 12/exon 13 junction of nucleotides 1411 and 1412 of SEQ ID NO:4; the exon 13/exon 14 junction of nucleotides 1491 and 1492 of SEQ ID NO:4; the exon 14/exon 15 junction of

nucleotides 1604 and 1605 of SEQ ID NO:4; the exon 15/exon 16 junction of nucleotides 1788 and 1789 of SEQ ID NO:4; the exon 16/exon 17 junction of nucleotides 1873 and 1874 of SEQ ID NO:4; the exon 17/exon 18 junction of nucleotides 2019 and 2020 of SEQ ID NO:4; the exon 18/exon 19 junction of nucleotides 2110 and 2111 of SEQ ID NO:4; the exon 19/exon 20 junction of nucleotides 2259 and 2260 of SEQ ID NO:4; the exon 20/exon 21 junction of nucleotides 2345 and 2346 of SEQ ID NO:4; the exon 21/exon 22 junction of nucleotides 2415 and 2416 of SEQ ID NO:4; and the exon 22/exon 23 junction of nucleotides 2534 and 2535 of SEQ ID NO:4. Junction nucleotides may be located at any position of the selected SEQ ID NO:4 fragment.

[0101] The present invention further relates to polynucleotides that hybridize to the above-described sequences. In this regard, the present invention especially relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 90% and preferably at least 95% identity and more preferably at least 97% identity between the sequences.

[0102] Furthermore, a major consideration associated with hybridization analysis of DNA or RNA sequences is the degree of relatedness the probe has with the sequences present in the specimen under study. This is important with a blotting technique (e.g., Southern or Northern Blot), since a moderate degree of sequence homology under nonstringent conditions of hybridization can yield a strong signal even though the probe and sequences in the sample represent non-homologous genes.

[0103] The particular hybridization technique is not essential to the invention, any technique commonly used in the art is within the scope of the present invention. Typical probe technology is described in United States Patent 4,358,535 to Falkow *et al.*,, incorporated by reference herein. For example, hybridization can be carried out in a solution containing 6 x SSC (10 x SSC: 1.5 M sodium chloride, 0.15 M sodium citrate, pH 7.0), 5 x Denhardt's (1 x Denhardt's: 0.2% bovine serum albumin, 0.2% polyvinylpyrrolidone, 0.02%

Ficoll 400), 10 mM EDTA, 0.5% SDS and about 10^7 cpm of nick-translated DNA for 16 hours at 65 °C. Additionally, if hybridization is to an immobilized nucleic acid, a washing step may be utilized wherein probe binding to polynucleotides of low homology, or nonspecific binding of the probe, may be removed. For example, a stringent wash step may involve a buffer of 0.2 x SSC and 0.5% SDS at a temperature of 65 °C.

[0104] Additional information related to hybridization technology and, more particularly, the stringency of hybridization and washing conditions may be found in Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1989), which is incorporated herein by reference.

[0105] Polynucleotides of the invention which are sufficiently identical to a nucleotide sequences contained in SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:83 or SEQ ID NO:84 or in the cDNA inserts of ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809, ATCC Deposit No. 326637, ATCC Deposit No. PTA-4611 or ATCC Deposit No. PTA-4610 may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding *de novo* DNA cytosine methyltransferase proteins and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the *de novo* DNA cytosine methyltransferase genes. Such hybridization techniques are known to those of skill in the art. Typically, these nucleotide sequences are at least about 90% identical, preferably at least about 95% identical, more preferably at least about 97%, 98% or 99% identical to that of the reference. The probes generally will comprise at least 15 nucleotides. Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

[0106] The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

[107] The present invention also provides isolated polynucleotides encoding a mouse Dnmt3a2 and human DNMT3A2 promoter regions as set forth in SEQ ID NO:118 and SEQ ID NO:119, respectively, that is capable of directing expression of mouse and human *de novo* cytosine methyltransferases. The present invention further provides a nucleic acid construct or vector, comprising a mouse Dnmt3a2 or human DNMT3A2 promoter having a nucleotide sequence of SEQ ID NO:118 or 119, respectively, or an operative fragment thereof having promoter activity, and host cells harboring the same.

[108] In some embodiments, the promoter sequence can be modified by the addition of sequences, such as enhancers, or deletions of nonessential and/or undesired sequences. The promoter sequences can be sufficiently similar to that of the native promoter to provide for the desired specificity of transcription of a DNA sequence of interest. The promoter sequences can include natural and synthetic sequences as well as sequences which may be a combination of synthetic and natural sequences.

[109] The present invention is further directed to isolated polynucleotides comprising promoter fragments of mouse Dnmt3a2. Such fragments include nucleotides 1-100, 1-80, 1-60, 1-35, 10-100, 20-100 and 40-100 of SEQ ID NO:118. Other fragments include nucleotides 1-722, 449-699, 460-660, 475-640, 485-620, 490-600, 500-590, 525-575, 449-690, 449-670, 449-630, 449-590, 449-550, 449-530, 460-699, 480-699, 510-699, 530-699, 550-699, 590-699, 620-699, 600-1150, 650-1100, 700-1050, 750-1050, 1530-1840, 1550-1800, 1550-1770, 1550-1760, 1550-1700, 1550-1680, 1550-1640, 1550-1600, 1575-1840, 1600-1840, 1620-1840, 1650-1840, 1700-1840, 1730-1840, 1770-1840, 1790-1840, 1500-2095, 1530-2095, 1570-2095, 1620-2095, 1650-2095, 1690-2095, 1720-2095, 1750-2095, 1790-2095, 1820-2095, 1900-2095, 2000-2095, 1500-2070, 1550-2025, 1550-2000, 1550-1975, 1550-1950, 1550-1940, 1550-1900, 1550-1870 and 1550-1830 of SEQ ID NO:118.

[110] The present invention further relates to isolated polynucleotides comprising promoter sequence fragments of human DNMT3A2. Such fragments

include nucleotides 1-100, 1-80, 1-60, 1-35, 10-100, 20-100 and 40-100 of SEQ ID NO:119. Other fragments include nucleotides 400-700, 450-690, 475-660, 485-640, 490-620, 500-600, 525-595, 400-690, 450-670, 450-630, 450-590, 450-550, 450-530, 450-699, 450-699, 500-700, 530-700, 550-700, 590-700, 620-700, 600-925, 650-875, 700-800, 750-800, 1280-1586, 1300-1550, 1300-1520, 1300-1490, 1300-1450, 1300-1420, 1300-1390, 1300-1350, 1325-1590, 1350-1580, 1370-1580, 1400-1580, 1440-1580, 1480-1580, 1520-1590,1540-1580, 1500-1850, 1530-1850, 1570-1850, 1620-1850, 1650-1850, 1690-1850, 1720-1850 1475-1530, 1480-1520, 1490-1520, 1495-1520, 1724-2065, 1740-2055, 1760-2070, 1770-2050, 1790-2035, 1800-2020, 1820-2000, 1825-1990, 1845-1980, 1860-1950, 1870-1920 and 1890-1910.

[111] In some embodiments, the invention provides isolated polynucleotides at least 50% identical, preferably 55%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to polynucleotide sequences encoding the Dnmt3a2 promoter sequence in SEQ ID NO:118 or 119, wherein the polynucleotide sequence has Dnmt3a2 promoter activity in embryonic stem cells.

[112] In other embodiments, the invention provides isolated polynucleotide sequence of SEQ ID NO:118, SEQ ID NO:119, or a fragment thereof that has promoter activity, operatively linked, in a transcriptional unit, to a DNA sequence encoding a protein of interest. In one embodiment, the DNA sequence encodes a protein of interest selected from the group consisting of SEQ ID NO:5, 6, 7, 8, 85, 86 and fragments thereof. In sone embodiments, the DNA sequence encodes a polypeptide fragment of SEQ ID NO:5, 6, 7, 8, 85 or 86 that possesses wild-type protein activity. In other embodiments, the DNA sequence encodes a polypeptide fragment of SEQ ID NO:5, 6, 7, 8, 85 or 86 that is a dominant negative mutant that inhibits endogenous *de novo* cytosine methyltransferase activity. In other embodiments, the DNA sequence operatively linked to the promoter sequences can be a reporter gene. The reporter gene can encode a fluorescent or light-emitting protein such as green fluorescent protein, yellow

fluorescent protein, blue fluorescent protein, phycobiliprotein, luciferase, or apoaequorin. In other embodiments, the reporter gene can encode B-galactosidase or chloramphenicol acetyltransferase.

[113] The promoter sequences as described herein are particularly useful for directing expression of operably linked genes in mammalian cells. In a preferred embodiment, the promoter sequences are used to direct expression of transgenes in stem cells. In other embodiments, the cells are embryonic cells. In another embodiment, the cells are cancer cells.

III. Vectors, Host Cells, and Recombinant Expression

[114] The present invention also relates to vectors that comprise a polynucleotide of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

[115] For recombinant production, host cells can be genetically engineered to incorporate expression systems for polynucleotides of the invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual,* 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). For example, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or any other means known in the art may be utilized.

[116] Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli, Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as

Drosophila S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

[0117] A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophages, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al.,, Molecular Cloning: A Laboratory Manual (supra).*

[0118] RNA vectors may also be utilized for the expression of the *de novo* DNA cytosine methyltransferases disclosed in this invention. These vectors are based on positive or negative strand RNA viruses that naturally replicate in a wide variety of eukaryotic cells (Bredenbeek, P.J. and Rice, C.M., Virology 3: 297-310, (1992)). Unlike retroviruses, these viruses lack an intermediate DNA life-cycle phase, existing entirely in RNA form. For example, alpha viruses are used as expression vectors for foreign proteins because they can be utilized in a broad range of host cells and provide a high level of expression; examples of viruses of this type include the Sindbis virus and Semliki Forest virus (Schlesinger, S., TIBTECH 11: 18-22, (1993); Frolov, I., *et al.,, Proc. Natl. Acad. Sci. (USA)* 93: 11371-11377, (1996)). As exemplified by Invitrogen's Sinbis expression system, the investigator may conveniently maintain the recombinant molecule in DNA form (pSinrep5 plasmid) in the laboratory, but propagation in RNA form is feasible as well. In the host cell used for expression, the vector containing the

gene of interest exists completely in RNA form and may be continuously propagated in that state if desired.

[0119] For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment appropriate secretion signals may be incorporated into the desired polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

[0120] As used herein, the term "operably linked," when used in the context of a linkage between a structural gene and an expression control sequence, *e.g.*, a promoter, refers to the position and orientation of the expression control sequence relative to the structural gene so as to permit expression of the structural gene in any host cell. For example, an operable linkage would maintain proper reading frame and would not introduce any in frame stop codons.

[0121] As used herein, the term "heterologous promoter," refers to a promoter not normally and naturally associated with the structural gene to be expressed. For example, in the context of expression of a *de novo* DNA cytosine methyltransferase polypeptide, a heterologous promoter would be any promoter other than an endogenous promoter associated with the *de novo* DNA cytosine methyltransferase gene in non-recombinant mouse or human chromosomes. In specific embodiments of this invention, the heterologous promoter is a prokaryotic or bacteriophage promoter, such as the lac promoter, T3 promoter, or T7 promoter. In other embodiments, the heterologous promoter is a eukaryotic promoter.

[0122] In other embodiments, this invention provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter. As used herein, the term "a *de novo* DNA cytosine methyltransferase structural gene" refers to a nucleotide sequence at least about 90% identical to one of the following nucleotide sequences:

(a) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:85 or SEQ ID NO:86;

(b) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence encoded by the cDNA insert of ATCC Deposit No. 209933, ATCC Deposit No. 209934, ATCC Deposit No. 98809, ATCC Deposit No. PTA-4611, or ATCC Deposit No. PTA-4610; or

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

[0123] In preferred embodiments, the *de novo* DNA cytosine methyltransferase structural gene is 90%, and more preferably 91%, 92%, 93%, 94%, 95%, 97%, 98%, 99%, or 100% identical to one or more of nucleotide sequences (a), (b), or (c) *supra*.

[0124] In another embodiment the term "a *de novo* DNA cytosine methyltransferase structural gene" refers to a nucleotide sequence about 90% to 99% identical to one of the following nucleotide sequences:

(a) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence in SEQ ID NO:8;

(b) a nucleotide sequence encoding the *de novo* DNA cytosine methyltransferase polypeptide having the complete amino acid sequence encoded by the cDNA insert of ATCC Deposit No. 326637; or

(c) a nucleotide sequence complementary to any of the nucleotide sequences in (a) or (b).

[0125] In preferred embodiments, the *de novo* DNA cytosine methyltransferase structural gene is 90%, and more preferably 91%, 92%, 93%, 94%, 95%, 97%, 98%, or 99% identical to SEQ ID NO:8, ATCC Deposit No. 326637 or polynucleotides complementary thereto.

[0126] This invention also provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule does not encode a fusion protein comprising the *de novo* DNA cytosine methyltransferase structural gene or a fragment thereof.

[0127] This invention further provides an isolated nucleic acid molecule comprising a *de novo* DNA cytosine methyltransferase structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule is capable of expressing a *de novo* DNA cytosine methyltransferase polypeptide when used to transform an appropriate host cell.

[0128] This invention also provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to a sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86 wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 3' untranslated region of SEQ ID NO:1 (nucleotides 2942-4191), SEQ ID NO:2 (nucleotides 2847-4174), SEQ ID NO:3 (nucleotides 3090-4397), SEQ ID NO:4 (nucleotides 2677-4127), SEQ ID NO:83 (nucleotides 2215-2318) or SEQ ID NO:84 (nucleotides 2274-2371) or a fragment of the 3' untranslated region greater than 25, 50, 75, 100, or 125 bp in length.

[0129] This invention further provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to a sequence encoding a *de novo* DNA cytosine methyltransferase polypeptide having the amino acid sequence of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86 wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 5' untranslated region of SEQ ID NO:1 (nucleotides 1-216), SEQ ID NO:2

(nucleotides 1-268), SEQ ID NO:3 (nucleotides 1-352), SEQ ID NO:4 (nucleotides 1-114), SEQ ID NO:83 (nucleotides 1-147) or SEQ ID NO:84 (nucleotides 1-216) or a fragment of the 5' untranslated region greater than 25, 35, 45, 55, 65, 75, 85, or 90 baseband processor/MAC.

[0130] Suitable known prokaryotic promoters for use in the production of proteins of the present invention include the *E. coli lac*I and *lac*Z promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), adenovirus promoter, *Herpes* virus promoter, and metallothionein promoters, such as the mouse metallothionein-I promoter and tissue and organ-specific promoters known in the art.

[0131] If the *de novo* DNA cytosine methyltransferase polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If *de novo* DNA cytosine methyltransferase polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

[0132] *De novo* DNA cytosine methyltransferase polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

IV. Polypeptides of the Invention

[0133] The *de novo* DNA cytosine methyltransferase polypeptides of the present invention include the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86 as well as polypeptides and fragments which have activity and have at least 90% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86, or the relevant portion and more preferably at least 96%, 97% or 98% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86, and still more preferably at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to the polypeptide of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86.

[0134] The polypeptides of the present invention are preferably provided in an isolated form.

[0135] The polypeptides of the present invention include the polypeptide encoded by the deposited cDNAs; a polypeptide comprising amino acids from about 1 to about 908 in SEQ ID NO:5; a polypeptide comprising amino acids from about 1 to about 859 in SEQ ID NO:6; a polypeptide comprising amino acids from about 1 to about 912 in SEQ ID NO:7, a polypeptide comprising amino acids from about 1 to about 853 in SEQ ID NO:8, a polypeptide comprising amino acids from about 1 to about 689 in SEQ ID NO:85, and a polypeptide comprising amino acids from about 1 to about 689 in SEQ ID NO:86 as well as polypeptides which are at least about 90% identical, and more preferably at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the polypeptides described above and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

[0136] Polypeptides of the invention also include alternative splicing variants of the Dnmt3 sequences disclosed herein. For example, alternative variant spliced proteins of mouse Dnmt3b include but are not limited to a polypeptide wherein,

except for at least one conservative amino acid substitution, said polypeptide has a sequence selected from the group consisting of: (1) amino acid residues 1 to 362 and 383 to 859 from SEQ ID NO:2; and (2) amino acid residues 1 to 362 and 383 to 749 and 813 to 859 from SEQ ID NO:2; and alternative variant spliced proteins of human DNMT3B include but are not limited to a polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has a sequence selected from the group consisting of: (1) amino acid residues 1 to 355 and 376 to 853 from SEQ ID NO:4; and (2) amino acid residues 1 to 355 and 376 to 743 and 807 to 853 from SEQ ID NO:4.

[0137] Polypeptides of the invention also include isoforms of mouse Dnmt3a and human DNMT3A disclosed herein which may arise through the use of an alternative promoter of the Dnmt3a or DNMT3A gene. For example, isoforms of mouse Dnmt3a arising through differential promoter usage include but are not limited to a polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has the sequence encoded by SEQ ID NO:84. Isoforms of human DNMT3A arising through differential promoter usage include but are not limited to a polypeptide wherein, except for at least one conservative amino acid substitution, said polypeptide has the sequence encoded by SEQ ID NO:85.

[0138] The *de novo* DNA cytosine methyltransferase polypeptides may be a part of a larger protein such as a fusion protein. It is often advantageous to include additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or additional sequence for stability during recombinant production.

[0139] Biologically active fragments of the *de novo* DNA cytosine methyltransferase polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of one of the aforementioned *de novo* DNA cytosine methyltransferase polypeptides. As with *de novo* DNA cytosine methyltransferase polypeptides, fragments may be "free-standing," or comprised

within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. In the context of this invention, a fragment may constitute from about 10 contiguous amino acids identified in SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85, or SEQ ID NO:86. More specifically, polypeptide fragment lengths may be defined algebraically as follows: (a) for SEQ ID NO:5, as 10 + N, wherein N equals zero or any positive integer up to 898; (b) for SEQ ID NO:6, as 10 + N, wherein N equals zero or any positive integer up to 849; (c) for SEQ ID NO:7, as 10 + N, wherein N equals zero or any positive integer up to 902; (d) for SEQ ID NO:8, as 10 + N, wherein N equals zero or any positive integer up to 843; (e) for SEQ ID NO:85, as 10 + N, wherein N equals zero or any positive integer up to 679; and (f) for SEQ ID NO:86, as 10 + N, wherein N equals zero or any positive integer up to 679.

[0140] Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of *de novo* DNA cytosine methyltransferase polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Biologically active fragments are those that mediate protein activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

[0141] In a specific embodiment, the polypeptide fragments are SEQ ID NO:85 and SEQ ID NO:86.

[0142] Thus, the polypeptides of the invention include polypeptides having an amino acid sequence at least 90% identical to that of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86 or fragments thereof with at least 90% identity to the corresponding fragment of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7 or SEQ ID NO:8, SEQ ID NO:85 or SEQ ID NO:86, all of which retain the biological activity of the *de novo* DNA cytosine methyltransferase protein, including antigenic activity. Included in this group are variants of the defined sequence and fragment. Preferred variants are those that vary from the reference by conservative amino acid substitutions, *i.e.*, those that substitute a residue with another of like characteristics. Typical substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg, or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5 to 10, 1 to 5, or 1 to 2 amino acids are substituted, deleted, or added in any combination.

[0143] The *de novo* DNA cytosine methyltransferase polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

V. In Vitro DNA Methylation

[0144] One preferred embodiment of the invention enables the *in vitro* methylation at the C5 position of cytosine in DNA. The starting substrate DNA may be hemimethylated (i.e., one strand of the duplex DNA is methylated) or may lack methylation completely. The polypeptides of the invention, being *de novo* DNA cytosine methyltransferases, are uniquely suited to the latter function,

owing to the fact that, unlike maintenance methyltransferases, their preferred substrate is not hemimethylated DNA.

[0145] As exemplified in Examples 4 and 5, isolated polypeptides of the invention function as *in vitro* DNA methyltransferases when combined in an appropriately buffered solution with the appropriate cofactors and a substrate DNA. The substrate DNA may be selected from any natural source, e.g., genomic DNA, or a recombinant source such as a DNA fragment amplified by the polymerase chain reaction. The substrate DNA may be prokaryotic or eukaryotic DNA. In a preferred embodiment, the substrate DNA is mammalian DNA, and most preferredly, the substrate DNA is human DNA.

[0146] It will be well appreciated by those in the art that *in vitro* methylation of DNA may be used to direct or regulate the expression of said DNA in a biological system. For example, over-expression, under-expression or lack of expression of a particular native DNA sequence in a host cell or organism may be attributed to the fact that the DNA is under-methylated (hypomethylated) or not methylated. Thus, *in vitro* methylation of a recombinant form of said DNA, and the subsequent introduction of the methylated, recombinant DNA into the cell or organism, may effect an increase or decrease in the expression of the encoded polypeptide.

[0147] Also, it will be readily apparent to the skilled artisan that the *in vitro* methylation pattern will be maintained after introduction into a biological system by the action of maintenance methyltransferase polypeptides in said system.

[0148] In one embodiment of the invention, the biological system selected for the introduction of *in vitro* methylated DNA may be prokaryotic or eukaryotic. In a preferred embodiment, the biological system is mammalian, and the most preferred embodiment is when the biological system is human.

[0149] Methods for introducing the *in vitro* methylated DNA into the biological system are well known in the art, and the skilled artisan will recognize that the *in vitro* methylation of DNA may be a preliminary step to any system of gene therapy detailed herein.

VI. Genetic Screening and Diagnostic Assays

[0150] To map the human chromosome locations, the GenBank STS database was searched using Dnmt3a and Dnmt3b sequences as queries. The search identified markers WI-6283 (GenBank Accession number G06200) and SHGC-15969 (GenBank Accession number G15302) as matching the cDNA sequence of Dnmt3a and Dnmt3b, respectively. WI-6283 has been mapped to 2p23 between D2S171 and D2S174 (48-50 cM) on the radiation hybrid map by Whitehead Institute/MIT Center for Genome Research. The corresponding mouse chromosome location is at 4.0 cM on chromosome 12. SHGC-15969 has been mapped to 20pl 1.2 between D20S184 and D20S106 (48-50 cM) by Stanford Human Genome Center. The corresponding mouse chromosome locus is at 84.0 cM on chromosome 2.

[0151] These data are valuable as markers to be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins, University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritence of physically adjacent genes).

[0152] The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

[0153] This invention also relates to the use of *de novo* DNA cytosine methyltransferase polynucleotides for use as diagnostic reagents. Detection of a mutated form of a *de novo* DNA cytosine methyltransferase gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of the mutated *de novo* DNA cytosine

methyltransferase. Individuals carrying mutations in one or more *de novo* DNA cytosine methyltransferase genes may be detected at the DNA level by a variety of techniques.

[0154] Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled *de novo* DNA cytosine methyltransferase nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, e.g., Myers, *et al.,, Science 230*:1242 (1985)). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton, *et al.,, Proc. Natl. Acad. Sci. USA 85*:4397-4401 (1985)).

[0155] The diagnostic assays offer a process for diagnosing or determining a susceptibility to neoplastic disorders through detection of mutations in one or more *de novo* DNA cytosine methyltransferase genes by the methods described.

[0156] In addition, neoplastic disorders may be diagnosed by methods that determine an abnormally decreased or increased level of *de novo* DNA cytosine methyltransferase polypeptide or *de novo* DNA cytosine methyltransferase mRNA in a sample derived from a subject. Decreased or increased expression may be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides; for example, RT-PCR, RNase protection, Northern blotting and other hybridization methods may be utilized. Assay techniques that may be used to determine the level of a protein, such as an

de novo DNA cytosine methyltransferase protein, in a sample derived from a host are well known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western blot analysis and ELISA assays.

[0157] Additionally, methods are provided for diagnosing or determining a susceptibility of an individual to neoplastic disorders, comprising (a) assaying the *de novo* DNA cytosine methyltransferase protein gene expression level in mammalian cells or body fluid; and (b) comparing said *de novo* DNA cytosine methyltransferase protein gene expression level with a standard *de novo* DNA cytosine methyltransferase protein gene expression level whereby an increase or decrease in said *de novo* DNA cytosine methyltransferase gene expression level over said standard is indicative of an increased or decreased susceptibility to a neoplastic disorder.

VII. De novo DNA Cytosine Methyltransferase Antibodies

[0158] The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them may also be used as immunogens to produce antibodies immunospecific for the *de novo* DNA cytosine methyltransferase polypeptides. By "immunospecific" is meant that the antibodies have affinities for the polypeptides of the invention that are substantially greater in their affinities for related polypeptides such as the analogous proteins of the prior art.

[0159] Antibodies generated against the *de novo* DNA cytosine methyltransferase polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor, *et al.,, Immunology Today 4*:72 (1983)) and the EBV-

hybridoma technique (Cole, *et al.,, Monoclonal Antibodies and Cancer Therapy*, pp. 77-96, Alan R. Liss, Inc., (1985)).

[0160] Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) may also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

[0161] The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

[0162] Antibodies against *de novo* DNA cytosine methyltransferase polypeptides may also be employed to treat neoplastic disorders, among others.

VIII. Agonist and Antagonist Screening

[0163] The *de novo* DNA cytosine methyltransferase polypeptides of the present invention may be employed in a screening process for compounds which bind one of the proteins and which activate (agonists) or inhibit activation of (antagonists) one of the polypeptides of the present invention. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics (see Coligan, *et al.,, Current Protocols in Immunology 1*(2):Chapter 5 (1991)).

[0164] By "agonist" is intended naturally occurring and synthetic compounds capable of enhancing a *de novo* DNA cytosine methyltransferase activity (e.g., increasing the rate of DNA methylation). By "antagonist" is intended naturally occurring and synthetic compounds capable of inhibiting a *de novo* DNA cytosine methyltransferase activity.

[0165] DNA methylation is an important, fundamental regulatory mechanism for gene expression, and, therefore, the methylated state of a particular DNA

sequence may be associated with many pathologies. Accordingly, it is desirous to find both compounds and drugs which stimulate *de novo* DNA cytosine methyltransferase activity and which can inhibit the function of *de novo* DNA cytosine methyltransferase protein. In general, agonists are employed for therapeutic and prophylactic purposes including the treatment of ceratin types of neoplastic disorders. For example, *de novo* methylation of growth regulatory genes in somatic tissues is associated with tumorigenesis in humans (Laird, P. W. and Jaenisch, R. *Ann. Rev. Genet. 30*:441-464 (1996); Baylin, S. B. *et al.,, Adv. Cancer. Res. 72*:141-196 (1998); and Jones, P. A. and Gonzalgo, M. L. *Proc. Natl. Acad. Sci. USA 94*:2103-2105 (1997)).

[0166] In general, such screening procedures involve producing appropriate cells which express the polypeptide of the present invention. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli.* Cells expressing the protein (or cell membrane containing the expressed protein) are then contacted with a test compound to observe binding, stimulation or inhibition of a functional response.

[0167] Alternatively, the screening procedure may be an *in vitro* procedure in which the activity of isolated DNMT3 protein is tested in the presence of a potential agonist or antagonist of DNMT3 *de novo* DNA cytosine methyltransferase activity. Such in vitro assays are known to those skilled in the art, and by way of example are demonstrated in Example 4 and 5.

[0168] The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the protein is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving competition with a labeled competitor. Further, these assays may test whether the candidate compound affects activity of the protein, using detection systems appropriate to the cells bearing the protein at their surfaces. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the candidate compound is observed. Standard methods for conducting such screening assays are well understood in the art.

[0169] Examples of potential *de novo* DNA cytosine methyltransferase protein antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the substrate of the *de novo* DNA cytosine methyltransferase protein, e.g., small molecules which bind to the protein so that the activity of the protein is prevented.

IX. Gene Therapy Applications

[0170] For overview of gene therapy, see Strachan, T. & Read A.P., Chapter 20, "Gene Therapy and Other Molecular Genetic-based Therapeutic Approaches," (and references cited therein) in *Human Molecular Genetics*, BIOS Scientific Publishers Ltd. (1996).

[0171] Initial research in the area of gene therapy focused on a few well-characterized and highly publicized disorders: cystic fibrosis (Drumm, M.L. *et al.*, *Cell 62*:1227-1233 (1990); Gregory, R.J. *et al.*, *Nature 347*:358-363 (1990); Rich, D.P. *et al.*, *Nature 347*:358-363 (1990)); and Gaucher disease (Sorge, J. *et al.*, *Proc. Natl. Acad. Sci. (USA) 84*:906-909 (1987); Fink, J.K. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:2334-2338 (1990)); and certain forms of hemophilia-Bontempo, F.A. *et al.*, *Blood 69*:1721-1724 (1987); Palmer, T.D. *et al.*, *Blood 73*:438-445 (1989); Axelrod, J.H. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:5173-5177 (1990); Armentano, D. *et al.*, *Proc. Natl. Acad. Sci. (USA) 87*:6141-6145 (1990)); and muscular dystrophy (Partridge, T.A. *et al.*, *Nature 337*:176-179 (1989); Law, P.K. *et al.*, *Lancet 336*:114-115 (1990); Morgan, J.E. *et al.*, *J. Cell Biol. 111*:2437-2449 (1990)).

[0172] More recently, the application of gene therapy in the treatment of a wider variety of disorders is progressing, for example: cancer (Runnebaum, I.B., *Anticancer Res.* 17(4B): 2887-2890, (1997)), heart disease (Rader, D.J., *Int. J. Clin. Lab. Res.* 27(1): 35-43, (1997); Malosky, S., *Curr. Opin. Cardiol.* 11(4): 361-368, (1996)), central nervous system disorders and injuries (Yang, K., *et al.*, *Neurotrauma J.* 14(5): 281-297, (1997); Zlokovic, B.V., *et al.*, *Neurosurgery*

40(4): 789-803, (1997); Zlokovic, B.V., *et al.*, *Neurosurgery* 40(4): 805-812, (1997)), vascular diseases (Clowes, A.W., *Thromb. Haemost.* 78(1): 605-610, 1997), muscle disorders (Douglas, J.T., *et al.*, *Neuromuscul. Disord.* 7(5): 284-298, (1997); Huard, J., *et al.*, *Neuromuscul. Disord.* 7(5): 299-313, (1997)), rheumatoid arthritis (Evans, C.H., *et al.*, *Curr. Opin. Rheumatol.* 8(3): 230-234, (1996)) and epithelial tissue disorders (Greenhalgh, D.A., *et al.*, *Invest Dermatol. J.* 103(5 Suppl.): 63S-93S, (1994)).

[0173] In a preferred approach, one or more isolated nucleic acid molecules of the invention are introduced into or administered to the animal. Such isolated nucleic acid molecules may be incorporated into a vector or virion suitable for introducing the nucleic acid molecules into the cells or tissues of the animal to be treated, to form a transfection vector. Techniques for the formation of vectors or virions comprising the *de novo* DNA cytosine methyltransferase-encoding nucleic acid molecules are well known in the art and are generally described in "Working Toward Human Gene Therapy," Chapter 28 in *Recombinant DNA, 2nd Ed.*, Watson, J.D. *et al.*, eds., New York: Scientific American Books, pp. 567-581 (1992). An overview of suitable vectors or virions is provided in an article by Wilson, J.M. (*Clin. Exp. Immunol.* 107(Suppl. 1): 31-32, (1997)). Such vectors are derived from viruses that contain RNA (Vile, R.G., *et al.*, *Br. Med Bull.* 51(1): 12-30, (1995)) or DNA (Ali M., *et al.*, *Gene Ther.* 1(6): 367-384, (1994)). Example vector systems utilized in the art include the following: retroviruses (Vile, R.G., supra.), adenoviruses (Brody, S.L. *et al.*, *Ann. N.Y. Acad. Sci.* 716: 90-101, (1994)), adenoviral/retroviral chimeras (Bilbao, G., *et al.*, *FASEB J.* 11(8): 624-634, (1997)), adeno-associated viruses (Flotte, T.R. and Carter, B.J., *Gene Ther.* 2(6): 357-362, (1995)), herpes simplex virus (Latchman, D.S., *Mol. Biotechnol.* 2(2): 179-195, (1994)), Parvovirus (Shaughnessy, E., *et al.*, *Semin Oncol.* 23(1): 159-171, (1996)) and reticuloendotheliosis virus (Donburg, R., *Gene Therap.* 2(5): 301-310, (1995)). Also of interest in the art, the development of extrachromosomal replicating vectors for gene therapy (Calos, M.P., *Trends Genet.* 12(11): 463-466, (1996)).

[0174] Other, nonviral methods for gene transfer known in the art (Abdallah, B. *et al.*, *Biol. Cell* 85(1): 1-7, (1995)) might be utilized for the introduction of *de novo* DNA cytosine methyltransferase polynucleotides into target cells; for example, receptor-mediated DNA delivery (Philips, S.C., *Biologicals* 23(1): 13-16, (1995)) and lipidic vector systems (Lee, R.J. and Huang, L., *Crit. Rev. Ther. Drug Carrier Syst.* 14(2): 173-206, (1997)) are promising alternatives to viral-based delivery systems.

[0175] General methods for construction of gene therapy vectors and the introduction thereof into affected animals for therapeutic purposes may be obtained in the above-referenced publications, the disclosures of which are specifically incorporated herein by reference in their entirety. In one such general method, vectors comprising the isolated polynucleotides of the present invention are directly introduced into target cells or tissues of the affected animal, preferably by injection, inhalation, ingestion or introduction into a mucous membrane via solution; such an approach is generally referred to as "*in vivo*" gene therapy. Alternatively, cells, tissues or organs may be removed from the affected animal and placed into culture according to methods that are well-known to one of ordinary skill in the art; the vectors comprising the *de novo* DNA cytosine methyltransferase polynucleotides may then be introduced into these cells or tissues by any of the methods described generally above for introducing isolated polynucleotides into a cell or tissue, and, after a sufficient amount of time to allow incorporation of the *de novo* DNA cytosine methyltransferase polynucleotides, the cells or tissues may then be re-inserted into the affected animal. Since the introduction of a *de novo* DNA cytosine methyltransferase gene is performed outside of the body of the affected animal, this approach is generally referred to as "*ex vivo*" gene therapy.

[0176] For both *in vivo* and *ex vivo* gene therapy, the isolated *de novo* DNA cytosine methyltransferase polynucleotides of the invention may alternatively be operatively linked to a regulatory DNA sequence, which may be a *de novo* DNA cytosine methyltransferase promoter or an enhancer, or a heterologous regulatory

DNA sequence such as a promoter or enhancer derived from a different gene, cell or organism, to form a genetic construct as described above. This genetic construct may then be inserted into a vector, which is then used in a gene therapy protocol. The need for transcriptionally targeted and regulatable vectors providing cell-type specific and inducible promoters is well recognized in the art (Miller, N. and Whelan, J., *Hum. Gene Therap.* 8(7): 803-815, (1997); and Walther, W. and Stein, U., *Mol. Med. J.*, 74(7): 379-392, (1996)), and for the purposes of *de novo* DNA cytosine methyltransferase gene therapy, is incorporated herein by reference.

[0177] The construct/vector may be introduced into the animal by an *in vivo* gene therapy approach, *e.g.*, by direct injection into the target tissue, or into the cells or tissues of the affected animal in an *ex vivo* approach. In another preferred embodiment, the genetic construct of the invention may be introduced into the cells or tissues of the animal, either *in vivo* or *ex vivo*, in a molecular conjugate with a virus (*e.g.*, an adenovirus or an adeno-associated virus) or viral components (*e.g.*, viral capsid proteins; *see* WO 93/07283). Alternatively, transfected host cells, which may be homologous or heterologous, may be encapsulated within a semi-permeable barrier device and implanted into the affected animal, allowing passage of *de novo* DNA cytosine methyltransferase polypeptides into the tissues and circulation of the animal but preventing contact between the animal's immune system and the transfected cells (*see* WO 93/09222). These approaches result in increased production of *de novo* DNA cytosine methyltransferase by the treated animal via (a) random insertion of the *de novo* DNA cytosine methyltransferase gene into the host cell genome; or (b) incorporation of the *de novo* DNA cytosine methyltransferase gene into the nucleus of the cells where it may exist as an extrachromosomal genetic element. General descriptions of such methods and approaches to gene therapy may be found, for example, in U.S. Patent No. 5,578,461, WO 94/12650 and WO 93/09222.

[0178] Antisense oligonucleotides have been described as naturally occurring biological inhibitors of gene expression in both prokaryotes (Mizuno *et al.,*, *Proc. Natl. Acad. Sci. USA 81*:1966-1970 (1984)) and eukaryotes (Heywood, *Nucleic Acids Res. 14*:6771-6772 (1986)), and these sequences presumably function by hybridizing to complementary mRNA sequences, resulting in hybridization arrest of translation (Paterson, *et al.,*, *Proc. Natl. Acad. Sci. USA, 74*:4370-4374 (1987)).

[0179] Thus, another gene therapy approach utilizes antisense technology. Antisense oligonucleotides are short synthetic DNA or RNA nucleotide molecules formulated to be complementary to a specific gene or RNA message. Through the binding of these oligomers to a target DNA or mRNA sequence, transcription or translation of the gene can be selectively blocked and the disease process generated by that gene can be halted (*see*, for example, Jack Cohen, *Oligodeoxynucleotides, Antisense Inhibitors of Gene Expression*, CRC Press (1989)). The cytoplasmic location of mRNA provides a target considered to be readily accessible to antisense oligodeoxynucleotides entering the cell; hence much of the work in the field has focused on RNA as a target. Currently, the use of antisense oligodeoxynucleotides provides a useful tool for exploring regulation of gene expression *in vitro* and in tissue culture (Rothenberg, *et al.,*, *J. Natl. Cancer Inst. 81*:1539-1544 (1989)).

[0180] Antisense therapy is the administration of exogenous oligonucleotides which bind to a target polynucleotide located within the cells. For example, antisense oligonucleotides may be administered systemically for anticancer therapy (Smith, International Application Publication No. WO 90/09180).

[0181] The antisense oligonucleotides of the present invention include derivatives such as S-oligonucleotides (phosphorothioate derivatives or S-oligos, *see*, Jack Cohen, *supra*). S-oligos (nucleoside phosphorothioates) are isoelectronic analogs of an oligonucleotide (O-oligo) in which a nonbridging oxygen atom of the phosphate group is replaced by a sulfur atom. The S-oligos of the present invention may be prepared by treatment of the corresponding O-

oligos with 3*H*-1,2-benzodithiol-3-one-1,1-dioxide which is a sulfur transfer reagent. *See* Iyer *et al.,, J. Org. Chem. 55*:4693-4698 (1990); and Iyer *et al.,, J. Am. Chem. Soc. 112*:1253-1254 (1990), the disclosures of which are fully incorporated by reference herein.

[0182] As described herein, sequence analysis of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:83, or the SEQ ID NO:84 cDNA clone shows that sequence that is nonhomologous to known DNA methyltransferase sequences may be identified (*see* Figures 1 and 4). Thus, the antisense oligonucleotides of the present invention may be RNA or DNA that is complementary to and stably hybridize with such sequences that are specific for a *de novo* DNA cytosine methyltransferase gene of the invention. Use of an oligonucleotide complementary to such regions allows for selective hybridization to a *de novo* DNA cytosine methyltransferase mRNA and not to an mRNA encoding a maintenance methyltransferase protein.

[0183] Preferably, the antisense oligonucleotides of the present invention are a 15 to 30-mer fragment of the antisense DNA molecule coding for unique sequences of the *de novo* DNA cytosine methyltransferase cDNAs. Preferred antisense oligonucleotides bind to the 5′-end of the *de novo* DNA cytosine methyltransferase mRNAs. Such antisense oligonucleotides may be used to down regulate or inhibit expression of the gene.

[0184] Other criteria that are known in the art may be used to select the antisense oligonucleotides, varying the length or the annealing position in the targeted sequence.

[0185] Included as well in the present invention are pharmaceutical compositions comprising an effective amount of at least one of the antisense oligonucleotides of the invention in combination with a pharmaceutically acceptable carrier. In one embodiment, a single antisense oligonucleotide is utilized.

[0186] In another embodiment, two antisense oligonucleotides are utilized which are complementary to adjacent regions of the genome. Administration of two antisense oligonucleotides that are complementary to adjacent regions of the

genome or corresponding mRNA may allow for more efficient inhibition of genomic transcription or mRNA translation, resulting in more effective inhibition of protein or mRNA production.

[0187] Preferably, the antisense oligonucleotide is coadministered with an agent which enhances the uptake of the antisense molecule by the cells. For example, the antisense oligonucleotide may be combined with a lipophilic cationic compound which may be in the form of liposomes. The use of liposomes to introduce nucleotides into cells is taught, for example, in U.S. Patent Nos. 4,897,355 and 4,394,448, the disclosures of which are incorporated by reference in their entirety (*see also* U.S. Patent Nos. 4,235,871, 4,231,877, 4,224,179, 4,753,788, 4,673,567, 4,247,411, and 4,814,270 for general methods of preparing liposomes comprising biological materials).

[0188] Alternatively, the antisense oligonucleotide may be combined with a lipophilic carrier such as any one of a number of sterols including cholesterol, cholate and deoxycholic acid. A preferred sterol is cholesterol.

[0189] In addition, the antisense oligonucleotide may be conjugated to a peptide that is ingested by cells. Examples of useful peptides include peptide hormones, antigens or antibodies, and peptide toxins. By choosing a peptide that is selectively taken up by the targeted tissue or cells, specific delivery of the antisense agent may be effected. The antisense oligonucleotide may be covalently bound via the 5'OH group by formation of an activated aminoalkyl derivative. The peptide of choice may then be covalently attached to the activated antisense oligonucleotide via an amino and sulfhydryl reactive hetero bifunctional reagent. The latter is bound to a cysteine residue present in the peptide. Upon exposure of cells to the antisense oligonucleotide bound to the peptide, the peptidyl antisense agent is endocytosed and the antisense oligonucleotide binds to the target mRNA to inhibit translation (Haralambid *et al.*,, WO 8903849 and Lebleu *et al.*,, EP 0263740).

[0190] The antisense oligonucleotides and the pharmaceutical compositions of the present invention may be administered by any means that achieve their

intended purpose. For example, administration may be by parenteral, subcutaneous, intravenous, intramuscular, intraperitoneal, or transdermal routes. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired.

[0191] Compositions within the scope of this invention include all compositions wherein the antisense oligonucleotide is contained in an amount effective to achieve the desired effect, for example, inhibition of proliferation and/or stimulation of differentiation of the subject cancer cells. While individual needs vary, determination of optimal ranges of effective amounts of each component is with the skill of the art.

[0192] Alternatively, antisense oligonucleotides can be prepared which are designed to interfere with transcription of the gene by binding transcribed regions of duplex DNA (including introns, exons, or both) and forming triple helices (e.g., *see* Froehler *et al.*,, WO 91/06626 or Toole, WO 92/10590). Preferred oligonucleotides for triple helix formation are oligonucleotides which have inverted polarities for at least two regions of the oligonucleotide (*Id.*). Such oligonucleotides comprise tandem sequences of opposite polarity such as 3′---5′-L-5′---3′, or 5′---3′-L-3′---5′, wherein L represents a 0-10 base oligonucleotide linkage between oligonucleotides. The inverted polarity form stabilizes single-stranded oligonucleotides to exonuclease degradation (Froehler *et al.*,, *supra*). The criteria for selecting such inverted polarity oligonucleotides is known in the art, and such preferred triple helix-forming oligonucleotides of the invention are based upon SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:83 or SEQ ID NO:84.

[0193] In therapeutic application, the triple helix-forming oligonucleotides can be formulated in pharmaceutical preparations for a variety of modes of administration, including systemic or localized administration, as described above.

[0194] The antisense oligonucleotides of the present invention may be prepared according to any of the methods that are well known to those of ordinary skill in the art, as described above.

[0195] Another gene therapy approach that may be utilized to alter expression of the *de novo* DNA methyl transferase genes of the invention is RNA interference (RNAi). The ability to specifically inhibit gene function in a variety of organisms utilizing double-stranded RNA (dsRNA)-mediated interference is well known in the fields of molecular biology (see for example C. P. Hunter, Current Biology 9:R440-442 (1999); Hamilton *et al.*, Science, 286:950-952 (1999); and S. W. Ding, Current Opinions in Biotechnology 11:152-156 (2000) hereby incorporated by reference in their entireties). Double-stranded RNA (dsRNA) that is homologous to a gene (or fragment therof) of interest is introduced into cells and effectively blocks expression of that gene in cells. The dsRNA molecules are digested *in vivo* to 21-23 nt fragment small interfering RNAs (siRNAs) which mediate the RNAi effect. In *C. elegans* and *Drosophila*, RNAi is induced by delivery of long dsRNA (up to 1-2 kb) produced by *in vitro* transcription. In mammalian cells, introduction of long dsRNA elicits a strong antiviral response that blocks any gene-specific silencing. However, introduction of 21 nt siRNAs with 2 nt 3' overhangs into mammalian cells does not stimulate the antiviral response and effectively targets specific mRNAs for gene silencing. The specificity of this gene silencing mechanism is extremely high, blocking expression only of targeted genes, while leaving other genes unaffected. Expression of *de novo* DNA methyl transferase transcripts of the invention may be turned off, for example, by delivery of siRNAs or vectors encoding the same into gonads or early embryos. In another embodiment, the siRNAs are delivered to cells or tissues to turn off expression of one or more *De novo* DNA methyl transferases. In a preferred embodiment, the cells are cancer cells. The artisan will appreciate that the siRNAs may be delivered to cells using an *in vivo* or *ex vivo* approach. Prefered *ex vivo* approaches involve transferring siRNAs to blood cells, bone marrow-derived cells, or stem cells.

[0196] The siRNAs or vectors encoding the same may be delivered to cells by techniques known in the art as described above. Further, the siRNAs may be prepared by any methods that are known in the art, including, but not limited to, oligonucleotide synthesis, *in vitro* transcription, ribonuclease digestion, or generation of siRNAs *in vivo*. In one embodiment, the siRNAs may be produced from vectors that are introduced into cells. The vectors may be introduced by any known methods in the art, including but not limited to transfection, electroporation, or viral delivery systems. Preferred vectors are the p*Silencer* siRNA expression vectors, p*Silencer* 2.0-U6 and p*Silencer* 3.0-H1. In a further embodiment, transcription of the siRNAs is driven by a RNA polymerase III (pol III) promoter. The pol III promoter may be derived from any gene that is under the control of RNA polymerase III, including but not limited to H1 or U6.

[0197] The siRNAs of the invention are encoded by nucleotide sequences within SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:83 or SEQ ID NO:84. In one embodiment, the siRNAs are about 20-1000 nucleotides in length. In another embodiment, the siRNAs are about 20-500 nucleotides in length. In another embodiment, the siRNAs are about 20-100 nucleotides in length. In another embodiment, the siRNAs are about 20-50 nucleotides in length. In a preferred embodiment, the siRNAs are about 21-23 nucleotides in length. The siRNAs may be produced by PCR amplification of genomic DNA or cDNA, using primers derived from *de novo* DNA methyl transferase sequence, and cloned into expression vectors for siRNA production. In another embodiment, oligonucleotides that correspond to *de novo* DNA methyl transferase sequence may be chemically synthesized and inserted into expression vectors for siRNA production. The siRNAs or vectors encoding the same are introduced into cells to block expression of the *de novo* methyl transferase polypeptides. siRNA can also be produced by chemical synthesis of oligonucleotide of RNA of 21-23 nucleotides. In one embodiment, the *de novo* methyl transferase polypeptides are selected from the group consisting of mouse Dnmt3a, Dnmt3a2, Dnmt3b1, Dnmt3b2, Dnmt3b3, Dnmt3b4, Dnmt3b5,

Dnmt3b6, and human DNMT3A, DNMT3A2, DNMT3B1, DNMT3B2, DNMT3B3, DNMT3B4, DNMT3B5 and DNMT3B6.

[0198] In one embodiment, the siRNAs are composed of nucleotides A, G, T, C, or U. Additionally, the siRNAs may be composed of unusual or modified nucleotides including but not limited to inosinic acid, 1-methyl inosinic acid, 1-methyl guanylic acid, NN-dimethyl guanylic acid, pseudouridylic acid, ribothymidylic acid, 5-hydroxymethylcytosine, and 5-hydroxymethyluridine. RNA may be synthesized either *in vivo* or *in vitro* and later introduced into cells. Endogenous RNA polymerase of the cell may mediate transcription *in vivo*, or cloned RNA polymerase can be used for transcription *in vitro*. For transcription from a transgene *in vivo* or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the RNA strand (or strands); the promoters may be known inducible promoters that respond to infection, stress, temperature, wounding, or chemicals. Inhibition may be targeted by specific transcription in an organ, tissue, or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The RNA strands may or may not be polyadenylated; the RNA strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. RNA may be chemically or enzymatically synthesized by manual or automated reactions. The RNA may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and production of an expression construct are known in the art (see, for example, WO 97/32016; U.S. Pat. Nos. 5,593,874; 5,698,425; 5,712,135; 5,789,214; and 5,804,693; and the references cited therein). If synthesized chemically or by *in vitro* enzymatic synthesis, the RNA may be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the RNA may be used with no or a minimum of purification to avoid losses due to

sample processing. The RNA may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

[0199] RNA containing nucleotide sequence identical to a fragment of the *de novo* DNA methyl transferase sequences are preferred for inhibition; however, RNA sequences with insertions, deletions, and point mutations relative to the *de novo* DNA methyl transferase sequences of the invention can also be used for inhibition. Sequence identity may optimized by sequence comparison and alignment algorithms known in the art (see Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Alternatively, the duplex region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a fragment of the target gene transcript.

[0200] Ribozymes provide an alternative method to inhibit mRNA function. Ribozymes may be RNA enzymes, self-splicing RNAs, and self-cleaving RNAs (Cech *et al.*,, *Journal of Biological Chemistry 267*:17479-17482 (1992)). It is possible to construct *de novo* ribozymes which have an endonuclease activity directed in *trans* to a certain target sequence. Since these ribozymes can act on various sequences, ribozymes can be designed for virtually any RNA substrate. Thus, ribozymes are very flexible tools for inhibiting the expression of specific genes and provide an alternative to antisense constructs.

[0201] A ribozyme against chloramphenicol acetyltransferase mRNA has been successfully constructed (Haseloff *et al., Nature 334*:585-591 (1988); Uhlenbeck *et al., Nature 328*:596-600 (1987)). The ribozyme contains three structural domains: 1) a highly conserved region of nucleotides which flank the cleavage site in the 5′ direction; 2) the highly conserved sequences contained in naturally occurring cleavage domains of ribozymes, forming a base-paired stem; and 3) the

regions which flank the cleavage site on both sides and ensure the exact arrangement of the ribozyme in relation to the cleavage site and the cohesion of the substrate and enzyme. RNA enzymes constructed according to this model have already proved suitable *in vitro* for the specific cleaving of RNA sequences (Haseloff *et al.,, supra*).

[0202] Alternatively, hairpin ribozymes may be used in which the active site is derived from the minus strand of the satellite RNA of tobacco ring spot virus (Hampel *et al.,, Biochemistry 28*:4929-4933 (1989)). Recently, a hairpin ribozyme was designed which cleaves human immunodeficiency virus type 1 RNA (Ojwang *et al.,, Proc. Natl. Acad. Sci. USA 89*:10802-10806 (1992)). Other self-cleaving RNA activities are associated with hepatitis delta virus (Kuo *et al.,, J. Virol. 62*:4429-4444 (1988)).

[0203] As discussed above, preferred targets for ribozymes are the *de novo* DNA cytosine methyltransferase nucleotide sequences that are not homologous with maintenance methyltransferase sequences such as Dnmt 1 or Dnmt 2. Preferably, the ribozyme molecule of the present invention is designed based upon the chloramphenicol acetyltransferase ribozyme or hairpin ribozymes, described above. Alternatively, ribozyme molecules are designed as described by Eckstein *et al.,* (International Publication No. WO 92/07065) who disclose catalytically active ribozyme constructions which have increased stability against chemical and enzymatic degradation, and thus are useful as therapeutic agents.

[0204] In an alternative approach, an external guide sequence (EGS) can be constructed for directing the endogenous ribozyme, RNase P, to intracellular mRNA, which is subsequently cleaved by the cellular ribozyme (Altman *et al.,*, U.S. Patent No. 5,168,053). Preferably, the EGS comprises a ten to fifteen nucleotide sequence complementary to an mRNA and a 3′-NCCA nucleotide sequence, wherein N is preferably a purine (*Id.*). After EGS molecules are delivered to cells, as described below, the molecules bind to the targeted mRNA species by forming base pairs between the mRNA and the complementary EGS

sequences, thus promoting cleavage of mRNA by RNase P at the nucleotide at the 5′side of the base-paired region (*Id.*).

[0205] Included as well in the present invention are pharmaceutical compositions comprising an effective amount of at least one ribozyme or EGS of the invention in combination with a pharmaceutically acceptable carrier. Preferably, the ribozyme or EGS is coadministered with an agent which enhances the uptake of the ribozyme or EGS molecule by the cells. For example, the ribozyme or EGS may be combined with a lipophilic cationic compound which may be in the form of liposomes, as described above. Alternatively, the ribozyme or EGS may be combined with a lipophilic carrier such as any one of a number of sterols including cholesterol, cholate and deoxycholic acid. A preferred sterol is cholesterol.

[0206] The ribozyme or EGS, and the pharmaceutical compositions of the present invention may be administered by any means that achieve their intended purpose. For example, administration may be by parenteral, subcutaneous, intravenous, intramuscular, intra-peritoneal, or transdermal routes. The dosage administered will be dependent upon the age, health, and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment, and the nature of the effect desired. For example, as much as 700 milligrams of antisense oligodeoxynucleotide has been administered intravenously to a patient over a course of 10 days (i.e., 0.05 mg/kg/hour) without signs of toxicity (Sterling, "Systemic Antisense Treatment Reported," *Genetic Engineering News 12*(12):1, 28 (1992)).

[0207] Compositions within the scope of this invention include all compositions wherein the ribozyme or EGS is contained in an amount which is effective to achieve inhibition of proliferation and/or stimulate differentiation of the subject cancer cells, or alleviate AD. While individual needs vary, determination of optimal ranges of effective amounts of each component is with the skill of the art.

[0208] In addition to administering the antisense oligonucleotides, ribozymes, or EGS as a raw chemical in solution, the therapeutic molecules may be

administered as part of a pharmaceutical preparation containing suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the antisense oligonucleotide, ribozyme, or EGS into preparations which can be used pharmaceutically.

[0209] Suitable formulations for parenteral administration include aqueous solutions of the antisense oligonucleotides, dsRNAs, ribozymes, EGS in water-soluble form, for example, water-soluble salts. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, and/or dextran. Optionally, the suspension may also contain stabilizers.

[0210] Alternatively, antisense RNA molecules, ribozymes, and EGS can be coded by DNA constructs which are administered in the form of virions, which are preferably incapable of replicating *in vivo* (*see*, for example, Taylor, WO 92/06693). For example, such DNA constructs may be administered using herpes-based viruses (Gage *et al.*,, U.S. Patent No. 5,082,670). Alternatively, antisense RNA sequences, ribozymes, and EGS can be coded by RNA constructs which are administered in the form of virions, such as retroviruses. The preparation of retroviral vectors is well known in the art (*see*, for example, Brown *et al.*,, "Retroviral Vectors," in *DNA Cloning: A Practical Approach*, Volume 3, IRL Press, Washington, D.C. (1987)).

[0211] Specificity for gene expression may be conferred by using appropriate cell-specific regulatory sequences, such as cell-specific enhancers and promoters. Such regulatory elements are known in the art, and their use enables therapies designed to target specific tissues, such as liver, lung, prostate, kidney, pancreas, etc., or cell populations, such as lymphocytes, neurons, mesenchymal, epithelial, muscle, etc.

[0212] In addition to the above noted methods for inhibiting the expression of the de novo methyltransferase genes of the invention, gene therapeutic applications may be employed to provide expression of the polypeptides of the invention.

[0213] The invention further provides methods of inhibiting de novo methylation in cells comprising expressing Dnmt3b3 and/or Dnmt3b6 in cells.

[0214] The present invention is further illustrated by the following Examples. These Examples are provided to aid in the understanding of the invention and are not to be construed as a limitation thereof.

EXAMPLES

EXAMPLE 1

Cloning and Sequence Analysis of the Mouse Dnmt3a and Dnmt3b and the Human DNMT3A and DNMT3B Genes and Polypeptides

[0215] In search of a mammalian *de novo* DNA methyltransferase, two independent approaches were undertaken, based on the assumption that an unknown mammalian DNA methyltransferase must contain the highly conserved cytosine methyltransferase motifs in the catalytic domain of known methyltransferases (Lauster, R. *et al.,, J. Mol. Biol. 206*:305-312 (1989) and Kumar, S. *et al.,, Nucl. Acids Res. 22*:1-10 (1994)). Our first approach, an RT/PCR-based screening using oligonucleotide primers corresponding to the conserved motifs of the known cytosine DNA methyltransferases, failed to detect any novel methyltransferase gene from *Dnmt1* null ES cells (data not shown). The second approach was a tblastn search of the dbEST database using full length bacterial cytosine methyltransferase sequences as queries.

[0216] A search of the dbEST database was performed with the tblastn program (Altschul, S. F. *et al.,, J. Mol. Biol. 215*:403-410 (1990)) using bacterial cytosine methyltransferases as queries. Candidate EST sequences were used one by one as queries to search the non-redundant protein sequence database in GenBank with the blastx program. This process would eliminate EST clones corresponding

to known genes (including known DNA methyltransferases) and those which show a higher similarity to other sequences than to DNA methyltransferases. Two EST clones (GenBank numbers W76111 and N88352) were found after the initial search. Two more EST sequences (fl2227 and T66356) were later found after a blastn search of dbEST with the EST sequence of W76111 as a query. Two of the EST clones (W76111 and T66356) were deposited by the I.M.A.G.E. Consortium (Lawrence Livermore National Laboratory, Livermore, CA) and obtained from American Type Culture Collection (Manassas, VA). Sequencing of these two cDNA clones revealed that they were partial cDNA clones with large open reading frames corresponding to two related genes. The translated amino acid sequences revealed the presence of the highly conserved motifs characteristic of DNA cytosine methyltransferases. The EST sequences were then used as probes for screening mouse E7.5 embryo and ES cell cDNA libraries and a human heart cDNA library (Clontech, CA).

[0217] In a screening of the dbEST database using 35 bacterial cytosine-5 DNA methyltransferase sequences as queries, eight EST clones were found to have the highest similarity but not to be identical to the known cytosine-5-DNA methyltransferase genes. Six of the eight EST sequences were deposited by the I.M.A.G.E. Consortium (Lawrence Livermore National Laboratory, Livermore, CA) and obtained from TIGR/ATCC (American Type Culture Collection, Manassas, VA). Sequencing of these 6 cDNA clones revealed that they were partial cDNA clones with large open reading frames corresponding to three novel genes. The translated amino acid sequences revealed the presence of the highly conserved motifs characteristic of DNA cytosine methyltransferases. The EST sequences were then used as probes for screening a mouse ES cell cDNA library, a mouse E11.5 embryonic cDNA library (Clontech, CA) and human heart cDNA library.

[0218] Human and mouse cDNA libraries were screened using EST sequences as probes. Sequencing analysis of several independent cDNA clones revealed that two homologous genes were present in both human and mouse. This was

further confirmed by Southern analysis of genomic DNA, intron/exon mapping and sequencing of genomic DNA (data not shown). The full length mouse cDNAs for each gene were assembled and complete sequencing revealed that both genes contained the highly conserved cytosine methyltransferase motifs and shared overall 51% of amino acid identity (76% identity in the catalytic domain) (Fig. 3). Since these two genes showed little sequence similarities to Dnmt1(Bestor, T. H. *et al.,, J. Mol. Biol. 203*:971-983 (1988) and Yen, R-W. C. *et al.,, Nucleic Acids Res. 20*:2287-2291 (1992)) and a recently cloned putative DNA methyltransferase gene, Dnmt2 (*see* Yoder, J. A. and Bestor, T. H. *Hum. Mol. Genet. 7*:279-284 (1998)) and Okano, M., Xie, S. and Li, E., (submitted)), beyond the conserved methyltransferase motifs in the catalytic domain, they were named Dnmt3a and Dnmt3b.

[0219] The full length Dnmt3a and Dnmt3b genes encode 908 and 859 amino acid polypeptides, termed Dnmt3a and Dnmt3b1, respectively. Nucleotide and amino acid sequences of each are presented in Figures 1A, 1B, 2A, and 2B. The Dnmt3b gene also produces through alternative splicing at least two shorter isoforms of 840 and 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3, respectively, (Fig. 4).

[0220] To obtain full length human cDNA, fetal heart and fetal testis cDNA libraries were screened using EST clones as probes. Sequencing analysis of several overlapping DNMT3A cDNA clones indicates that the DNMT3A gene encodes a polypeptide of 912 amino acid residues. DNMT3B cDNA clones were not detected in the fetal heart library, but several DNMT3B cDNA clones were obtained after screening the fetal testis library. PCR screening of large cDNA clones from 24 human tissues was also performed using the Human Rapid-Screen™ cDNA Library Panels (OriGene Technologies, MD). The largest cDNA clone contained a 4.2 kb insert from a small intestine cDNA library. Sequencing analysis of overlapping cDNA clones indicated that the deduced full length DNMT3B consists of 853 amino acid residues. Since in-frame stop codons are

found upstream of the ATG of both DNMT3A and DNMT3B, it is concluded that these cDNA clones encode full-length DNMT3A and DNMT3B proteins.

[0221] The full length human DNMT3A and DNMT3B cDNAs encode 912 and 853 amino acid polypeptides, termed DNMT3A and DNMT3B1, respectively. Nucleotide and polypeptide sequences are presented in Figures 1C, 1D, 2C and 2D, respectively. The DNMT3B gene also produces through alternative splicing at least two shorter isoforms, termed DNMT3B2 and DNMT3B3, respectively. DNMT3B2 comprises amino acid residues 1 to 355 and 376 to 853 of SEQ ID NO:4; and DNMT3B3 comprises amino acid residues 1 to 355 and 376 to 743 and 807 to 853 of SEQ ID NO:4.

[0222] Also identified through screening was a related zebrafish gene, termed Zmt-3, which from the EST database (GenBank number AF135438).

[0223] The GenBank STS database was used to map chromosome localization by using DNMT3A and DNMT3B sequences as queries. The results identified markers WI-6283 (GenBank Accession number G06200) and SHGC-15969 (GenBank Accession number G15302), which matched the cDNA sequence of DNMT3A and DNMT3B, respectively. WI-6283 has been mapped to 2p23 between D2S171 and D2S174 (48-50 cM) on the radiation hybrid map by Whitehead Institute/MIT Center for Genome Research. The corresponding mouse chromosome location is at 4.0 cM on chromosome 12. SHGC-15969 has been mapped to 20pl 1.2 between D20S184 and D20S106 (48-50 cM) by Stanford Human Genome Center. The corresponding mouse chromosome locus is at 84.0 cM on chromosome 2.

[0224] Taking the advantage of the newly identified DNMT3A and DNMT3B cDNA sequences, the human genomic sequence database was searched by BLAST. While human DNMT3A cDNA did not match any related genornic sequences in the database, a DNMT3B genomic YAC clone from GenBank (AL035071) was identified when DNMT3B cDNA sequences were used as queries.

[0225] The DNMT3B cDNA and the genomic DNA GenBank (AL035071) clone were used to map all exons using BESTFIT of the GCG program. As shown in Figure 4C, there are total 23 exons, spanning some 48 kb genomic DNA. The putative first exon is located within a CpG island where the promoter is probably located as predicted by the GENSCAN program (Whitehead/MIT Center for Genome Research).

[0226] Sequencing of various cDNA clones indicates that the human DNMT3B gene contains three alternatively spliced exons, exons 10, 21 and 22. Similar to the mouse gene, DNMT3B1 contains all 23 exons, whereas DNMT3B2 lacks exon 10 and DNMT3B3 lacks exons 10, 21 and 22. The nucleotide sequences at the exon/intron boundaries are shown in Figure 4D. The elucidation of human *DNMT3B* gene structure may facilitate analysis of DNMT3B mutations in certain cancers with characteristic hypomethylation of genomic: DNA (Narayan, A., *et al.,, Int. J. Cancer 77*:833-838 (1998); Qu, G., *et al.,, Mutan. Res. 423*:91-101 (1999)).

[0227] Figure 3A presents an alignment of mouse Dnmt3a and Dnmt3b polypeptide sequences that was accomplished using the GCG program. The vertical lines indicate amino acid identity, while the dots and the colons indicate similarities. Dots in amino acid sequences indicate gaps introduced to maximize alignment. The conserved Cys-rich region is shaded. The full length mouse Dnmt3a and Dnmt3b genes encode 908 and 859 amino acid polypeptides. Furthermore, the analysis reveals that both genes contained the highly conserved cytosine methyltransferase motifs and share overall 51% of amino acid identity (76% identity in the catalytic domain). The Dnmt3b gene also produces at least two shorter isoforms of 840 and 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3, respectively, through alternative splicing (Fig. 4).

[0228] Figure 3B presents a GCG program alignment using the of the protein sequences of human DNMT3A and DNMT3B1. Vertical lines represent identical amino acid residues, whereas dots represent conserved changes. Dots in amino acid sequences indicate gaps introduced to maximize alignment.

[0229] In Figure 4A, presents a schematic diagram of the overall protein structures for mouse *Dnmt*1, mouse *Dnmt*2, a putative methyltransferase, and the family of Dnmt3a and Dnmt3b(1-3) methyltransferases. *Dnmt*1, Dnmt3a and Dnmt3bs all have a putative N-terminal regulatory domain. The filled bars represent the five conserved methyltransferase motifs (I, IV, VI, IX, and X). The shaded boxes in Dnmt3a and Dnmt3bs represent the Cys-rich region that shows no sequence homology to the Cys-rich, Zn^{2+}-binding region of *Dnmt*1 polypeptide. Sites of alternative splicing at amino acid residues 362-383 and 749-813 in Dnmt3bs are indicated.

[0230] An analysis of the human DNMT3 proteins provides similar results as with the mouse Dnmt proteins. Figure 4B presents a similar schematic of the human DNMT3 proteins and zebrafish Znmt3 protein. The homology between differences between these DNMT3 proteins is indicated by the percentage of sequence identity when compared to DNMT3A.

[0231] In addition, the genomic organization of the human DNMT3B1 locus is presented in Figure 4C as possessing 23 exons (filled rectangles), a CpG island (dotted rectangle),a translation initiation codon (ATG) and a stop codon (TAG) in exons 2 and 23, respectively. Figure 4D presents the size of the exons and introns as well as sequences (uppercase for exons and lowercase for introns) at exon/intron boundaries.

[0232] In Figure 5, sequence analysis of the catalytic domain indicates that this new family of DNA methyltransferases contains conserved amino acid residues in each of the five highly conserved motifs, but significant differences are discernible when compared to the known consensus sequences.

[0233] Figure 5A presents an alignment by ClustalW 1.7 of the amino acid sequences of the five highly conserved motifs in eukaryotic methyltransferase genes. Amino acid residues which are conserved in five or more genes are highlighted. The Dnmt3 family methyltransferases are most closely related to a bacterial DNA methyltransferase (*M. Spr.*). Sequence comparison of the catalytic domain of all known eukaryotic DNA methyltransferases and most of the

bacterial cytosine methyltransferases used in the tblastn search indicates that this family of methyltransferases are distantly related to all the known eukaryotic DNA methyltransferases, including the *Dnmt* 1 polypeptide from vertebrate and plant (Bestor, T. H. *et al.,, J. Mol. Biol. 203*:971-983 (1988), Yen, R-W. C. *et al.,, Nucleic Acids Res. 20*:2287-2291 (1992) and Finnegan, E. J. and Dennis, E. S. *Nucleic Acids Res. 21*:2383-2388 (1993)); the human and mouse *Dnmt* 2 polypeptides (Yoder, J. A. and Bestor, T. H. *Hum. Mol. Genet. 7*:279-284 (1998), Okano, M., Xie, S. & Li, E., (submitted)); and masc1 from Ascobolus (Malagnac, F. *et al.,, Cell 91*:281-290 (1997)), indicating that the Dnmt3 gene family originated from a unique prokaryotic prototype DNA methyltransferase during evolution.

[0234] The cysteine-rich region located upstream of the catalytic domain was found to be conserved among all of the DNMT3 proteins (Fig. 5B). This Cysteine-rich region, however, is unrelated to the Cysteine-rich (or Zn^{2+}-binding) region of DNMT1 (Bestor, T.H., *et al.,, J. Mo. Biol. 203*:971-983 (1998); Bestor, T.H., *EMBO J. 11*:2611-2617 (1992)). Interestingly, the Cysteine-rich domain of DNMT3 proteins shares homology with a similar domain found in the X-linked ATRX gene of the SNF2/SWI family (Picketts, D.J., *et al.,, Hum. Mol. Genet. 5*:1899-1907 (1996)), raising the interesting possibility that this domain may mediate protein-protein or protein-DNA interactions.

[0235] The evolutionary relatedness of cytosine-5 methyltransferases as shown by a non-rooted phylogenic tree is presented in Figure 5C. Amino acid sequences from motif I to motif VI of bacterial and eukaryotic cytosine-5 methyltransferases were used for sequence alignment, and the alignment data was analyzed by ClustalW 1.7 under conditions excluding positions with gaps. Results were visualized utilizing Phlip version 3.3. Amino acid sequences from motif IX to motif X were also analyzed and provided similar results (data not shown). (Abbreviation Ath; *Arabidopsis thaliana*, Urc; sea urchin, Xen; *Xenopus laevis*).

EXAMPLE 2

Baculovirus-mediated Expression of Dnmt3a and Dnmt3b

[0236] To test whether the newly cloned Dnmt3 genes encode active DNA methyltransferases, the cDNAs of Dnmt3a, Dnmt3b1, Dnmt3b2, and Dnmt1 were overexpressed in insect cells using the baculovirus-mediated expression system (Clontech, CA).

[0237] To construct the Dnmt3a expression vector, pSX134, the *Xma I/Eco RI* fragment of Dnmt3a cDNA was first cloned into the *Nco I/Eco RI* sites of pET2 ld with the addition of an *Xma I/Nco I* adapter (SX165: 5'-CATGGGCAGCAGCCATCATCATCATCATCATGGGAATTCCATGCCC TCCAGCGGCC (SEQ ID NO: 87) and SX166: 5'-GGGCATGGAATT CCCATGATGATGATGATGATGGCTGCTGCC) (SEQ ID NO: 88) that produced pSX132His. pSX134 was obtained by cloning the *EcoR I/Xba I* fragment of pSX 132His into the *EcoR I/Xba I* sites of pBacPAK9. The Dnmt3b1 and Dnmt3b2 expression vectors, pSX153 and pSX154, were constructed by cloning *Eco RI* fragments of Dnmt3b1 and Dnmt3b2 cDNA into the *Eco RI* site of pBacPAK9, respectively. The Dnmt1 expression vector pSX148 was constructed by cloning the *Bgl I/Sac I* fragment of Dnmt1 cDNA into the *Bgl II/Sac I* sites of pBacPAK-His2 with the addition of a *Bgl I/Bgl II* adapter (SX180: 5'-GATCTATGCCAGCGCGA ACAGCTCCAGCCCGAGTGCCTGCGCTTGCCTCCC (SEQ ID NO: 89) and SX181: 5'- AGGCAAGCGCAGGCACTCGGGCTGGAGCTGTT CGCGCTGGCATA) (SEQ ID NO: 90).

[0238] pSX134 (Dnmt3a), pSX153 (Dnmt3b1), pSX153 (Dnmt3b2) and pSX148 (*Dnmt*1) were used to make the recombinant baculoviruses according to the procedures recommended by the manufacturer. T175 flasks were used for cell culture and virus infection. Sf21 host cells were grown in the SF-900 II SFM medium with 10% of the certified FBS (both from GIBCO, MD) and infected with the recombinant viruses 12-24 hours after the cells were split when they

reached 90-95% affluence. After 3 days, the infected insect cells were harvested and frozen in the liquid nitrogen for future use.

EXAMPLE 3

RNA Expression Analysis

[0239] ES cells were routinely cultured on a feeder layer of mouse embryonic fibroblasts in DMEM medium containing LIF (500 units/ml) and were differentiated as embryoid bodies in suspension culture as described (Lei, H., *et al.,, Development 122*:3195-3205 (1996)). Ten days after seeding, embryoid bodies were harvested for RNA preparation.

[0240] Total RNA was prepared from ES cells, ovary and testis tissue using the GTC-CsCl centrifugation method, fractionated on a formaldehyde denaturing 1% agarose gel by electrophoresis and transferred to a nylon membrane. PolyA+ RNA blots (2μg per lane) of mouse and human tissues were obtained from Clontech, CA. All blots were hybridized to random-primed cDNA probes in hybridization solution containing 50% formamide at 42 °C and washed with 0.2 X SSC, 0.1% SDS at 65 °C and exposed to X-ray film (Kodak).

[0241] Fig. 6A presents mouse polyA+ RNA blots of adult tissues (left) and embryos (right) probed with full length Dnmt3a, Dnmt3b and a control β-actin cDNA probe. Each lane contains 2 μg of polyA+ RNA. (Ht, Heart; Br, Brain; Sp, Spleen; Lu, Lung; Li, Liver; Mu, Skeletal Muscle; Ki, Kidney; Te, Testis; and embryos at gestation days 7 (E7), 11 (E11), 15 (E15), and 17 (E17). Fig. 6B is a mouse total RNA blot (10 μg per lane) of ES cell and adult organ RNA samples and Fig. 6C shows a mouse total RNA blot (20 μg per lane) of undifferentiated (Undiff.) and differentiated (Diff.) ES cells RNA hybridized to Dnmt3a, Dnmt3b or β-actin probes.

[0242] It has been shown that the maintenance methylation activity is constitutively present in proliferating cells, whereas the *de novo* methylation activity is highly regulated. Active *de novo* methylation has been shown to occur

primarily in ES cells (or embryonic carcinoma cells), early postimplantation embryos and primordial germ cells (Jähaner, D. and Jaenish, R., "DNA Methylation in Early Mammalian Development," In *DNA Methylation: Biochemistry and Biological Significance*, Razin, A. *et al.,, eds.,* Springer-Verlag (1984) pp. 189-219; Razin, A., and Cedar, H., "DNA Methylation and Embryogenesis," in *DNA Methylation: Molecular Biology and Biological Significance*, Jost., J. P. *et al.,, eds.,* Birkhäuser Verlag, Basel, Switzerland (1993) pp. 343-357; Chaillet, J. R. *et al.,, Cell 66*:77-83 (1991); and Li, E. "Role of DNA Methylation in Development," in *Genomic Imprinting: Frontiers in Molecular Biology*, Reik, W. and Sorani, A. *eds.,* IRL Press, Oxford (1997) pp. 1-20). The expression of both Dnmt3a and Dnmt3b in mouse embryos, adult tissues and ES cells was examined. The results indicate that two Dnmt3a transcripts, 9.5 kb and 4.2kb, are present in embryonic and adult tissue RNA. The 4.2 kb transcript, corresponding to the size of the full length cDNA, was expressed at very low levels in most tissues, except for the E11.5 embryo sample (Fig. 6A). A single 4.4 kb Dnmt3b transcript is detected in embryo and adult organ RNAs, with relatively high levels in testes and E11.5 embryo samples (Fig. 6A). Interestingly, both genes are expressed at much higher levels in ES cells than in adult tissues (Fig. 6B), and their expression decreased dramatically upon differentiation of ES cells in culture (Fig. 6C). In addition, Dnmt3a and Dnmt3b expression levels are unaltered in Dnmt1-deficient ES cells (Fig. 6C), suggesting that regulation of Dnmt3a and Dnmt3b expression is independent of Dnmt1.

[0243] These results suggest that both Dnmt3a and Dnmt3b are expressed specifically in ES cells and E11.5 embryo and/or testes. The expression in the E11.5 embryo and testes may correlate with the presence of developing or mature germ cells in these tissues. Therefore, the expression pattern of Dnmt3a and Dnmt3b appears to correlate well with *de novo* methylation activities in development.

[0244] For the RNA expression analysis of human DNMT3 genes, polyA+ RNA blots were hybridized using DNMT3A and DNMT3B cDNA fragments as probes.

Results indicate that DNMT3A RNA was expressed ubiquitously and was readily detected in most tissues examined at levels slightly lower than DNMT1 RNA (Fig. 9). Three major DNMT3A transcripts, approximately 4.0, 4.4, and 9.5 kb, were detected. The relative expression level of the transcripts appeared to vary from tissue to tissue. Transcripts of similar sizes were also detected in mouse tissues. Results utilizing DNMT3B cDNA probes indicate that transcripts of about 4.2 kb were expressed at much lower levels in most tissues, but could be readily detected in the testis, thyroid and bone marrow (Fig. 9). Sequence analyses of different cDNA clones indicate the presence of alternatively spliced transcripts, although the size differences between these transcripts are too small to be detected by Northern analysis.

[0245] Hypermethylation of tumor suppressor genes is a common epigenetic lesion found in tumor cells (Laird, P.W. & Jaenisch, R., Ann. Rev. Genet. 30:441-464 (1996); Baylin, S.B., *Adv. Cancer Res. 72*:141-196 (1998)). To investigate whether DNMT3A and DNMT38 am abnormally activated in tumor cells, DNMT3 RNA expression was analyzed in several tumor cell lines by Northern blot hybridization. Results demonstrated that DNMT3A was expressed at higher levels in most tumor cell lines examined. (Figure 10). As in the normal tissues, three different size transcripts were also detected in tumor cells. The ratio of these transcripts appeared to be variable in different tumor cell lines. DNMT3B expression was dramatically elevated in most tumor cell lines examined though it was expressed at very low levels in normal adult tissues (Figure 10). The expression levels of both DNMT3A and DNMT3B appear to be comparable and proportional to that of DNMT1.

[0246] The murine Dnmt3a and Dnmt3b genes are highly expressed in undifferentiated ES cells, consistent with their potential role in *de novo* methylation during early embryonic development. Additionally, both genes are highly expressed in early embryos. Differences in their expression patterns in adult tissues in both human and mice suggest that each gene may have a distinct function in somatic tissues and may methylate different genes or genomic

sequences. The elevated expression of DNMT3 genes in human tumor cell lines suggests that the DNMT3 enzyme may be responsible for *de novo* methylation of CpG islands in tumor suppressor genes during tumor formation.

EXAMPLE 4

Methyltransferase Activity Assay

[0247] In order to demonstrate DNA cytosine methyltransferase activity, the polypeptides of the invention were expressed and purified from recombinant host cells for use in *in vitro* assays.

[0248] Infected insect Sf21 cells and NIH3T3 cells were homogenized by ultrasonication in lysis solution (20 mM Tris-HCl, pH7.4, 10 mM EDTA, 500 mM NaCl, 10% glycerol, lmM DTT, lmM PMSF, 1 ug/ml leupeptin, 10 ug/ml TPCK, 10 ug/ml TLCK) and cleared by centrifugation at 100,000 g for 20 min.

[0249] The methyltransferase enzyme assay was carried out as described previously (Lei, H. *et al.,, Development 122*:3195-3205 (1996)). DNA substrates used in the assays include: poly (dI-dC), poly (dG-dC) (Pharmacia Biotech), lambda phage DNA (Sigma), pBluescriptIISK (Stratagene, CA), pMu3 plasmid, which contains tandem repeats of 535bp RsaI-RsaI fragment of MMLV LTR region in pUC9, and oligonucleotides. The oligonucleotide sequences utilized include:

#1, 5'-AGACMGGTGCCAGMGCAGCTGAGCMGGATC-3' (SEQ ID NO: 91), #2, 5'-GATCMGGCTCAGCTGMGCTGGCACMGGTCT-3' (SEQ ID NO: 92), #3, 5'-AGACCGGTGCCAGCGCAGCTGAGCCGGATC-3' (SEQ ID NO: 93) and #4, 5'-GATCCGGCTCAGCTGCGCTGGCACCGGTCT-3' (SEQ ID NO: 94) (M represents 5-methylcytosine).

[0250] These sequences are the same as described in a previous study (Pradhan, S. *et al.,, Nucleic Acids Res. 25*:4666-4673 (1997)). Oligonucleotides were synthesized and purified by polyacrylamide gel electrophoresis (PAGE). To make double strand oligonucleotides, equimolar amounts of the two

complimentary oligonucleotides were heated at 94°C for 10 min., mixed, incubated at 78°C for 1 hr and cooled down slowly at room temperature. The annealing products were quantified for the yield of double-stranded oligonucleotides (dsDNA) by PAGE and methylene blue staining. In all cases, the yield of dsDNA was higher than 95%. The dsDNA of #1 and #2 were used as 'fully' methylated substrates, dsDNA of #1 and #4 as the hemi-methylated substrates, and dsDNA of #3 and #4 as unmethylated substrates.

[0251] For Southern analysis of the methylation of retrovirus DNA, 2 ug of pMMLV8.3, an 8.3kb *Hind* III fragment of Moloney murine leukemia virus cDNA in pBluescriptIISK, was methylated *in vitro* for 15 hrs under the same reaction conditions described above except that 160 uM of cold SAM was used instead of ^{3}H-methyl SAM. Then, an equal volume of the solution containing 1% SDS, 400 mM NaCl, and 0.2 mg/ml Proteinase K was added, and the sample was incubated at 37°C for 1 hr. After phenol/chloroform extraction, DNA was precipitated with ethanol, dried and dissolved in TE buffer. This procedure was repeated 5 times. An aliquot of DNA was purified after the first, third and fifth reaction, digested with *Hpa* II or *Msp* I in combination with *Kpn* I for 16 hrs, separated on 1% agarose gels, blotted and hybridized to the pMu3 probe.

[0252] In a standard methyltransferase assay, enzyme activity was detected with protein extracts from Sf21 cells overexpressing Dnmt3a and Dnmt3b polypeptides. Similar to the results obtained with the Dnmt1 polypeptide, the overexpressed Dnmt3 proteins were able to methylate various native and synthetic DNA substrates, among which poly(dI-dC) consistently gave rise to the highest initial velocity (Fig.7a). An analysis of the methylation of *Hpa* II sites in retroviral DNA by these enzymes was also performed. An MMLV full length cDNA was methylated for 1-5 times by incubation with protein extract from control Sf21 cells or Sf21 cells infected with baculoviruses expressing *Dnmt*1, Dnmt3a or Dnmt3b polypeptides. The *Hpa* II/*Msp* I target sequence, CCGG, is resistant to the *Hpa* II restriction enzyme, but sensitive to *Msp* I digestion when the internal C is methylated, and the restriction site becomes resistant to *Msp* I

digestion when the external C is methylated (Jentsch, S. *et al.,, Nucleic Acids Res. 9*:2753-2759 (1981)). Both Dnmt3a and Dnmt3b polypeptides could methylate multiple *Hpa* II sites in the 3' LTR regions of the MMLV DNA, as indicated by the presence of *Hpa* II-resistant fragments, though less efficiently than *Dnmt*1 polypeptide (Fig. 7b). Significantly, even after five consecutive rounds of *in vitro* methylation, the viral DNA was completely digested by *Msp* I. This result indicates that both Dnmt3a and Dnmt3b polypeptides methylate predominantly the internal cytosine residues, therefore, CpGs. Previously it was shown that the same region of the proviral DNA was efficiently methylated in *Dnmt1* null ES cells infected by the MMLV virus (Lei, H. *et al.,, Development 122*:3195-3205 (1996)).

[0253] Fig. 7A shows ^{3}H-methyl incorporation into different DNA substrates (poly (dI-dC), poly (dG-dC) (squares), lambda phage DNA (circles), pBluescriptIISK (triangles), and pMu3 (diamonds)) when incubated with protein extracts of Sf21 cells expressing Dnmt1, Dnmt3a, or Dnmt3b1. Fig. 7B shows Southern blot analysis of the *in vitro* methylation of untreated pMMLV DNA (lanes 1-3) and pMMLV DNA incubated with MT1 (lane 4-10), MT3α (lanes 11-15), MT3β (lanes 16-20) or control Sf21 (lanes 21-25) extracts that were digested with *Kpn* I(K), *Kpn* I and *Msp* I (K/M) or *Kpn* I and *Hpa* II (K/H). Restriction enzyme digested samples were then subjected to Southern blot analysis using the pMu3 probe.

[0254] *Dnmt*1 protein appears to function primarily as a maintenance methyltransferase because of its strong preference for hemimethylated DNA and direct association with newly replicated DNA (Leonhardt, H. *et al.,, Cell 71*:865-873 (1992)). To determine whether Dnmt3a and Dnmt3b polypeptides show any preference for hemimethylated DNA over unmethylated DNA, a comparison was done to examine the methylation rate of unmethylated versus hemimethylated oligonucleotides. Gel-purified double stranded oligonucleotides were incubated with protein extracts of Sf21 cells expressing Dnmt1, Dnmt3a, Dnmt3bl, Dnmt3b2 or NIH3T3 cell extract (unmethylated substrates (open circles), hemi-

methylated substrates (half black diamonds) or completely methylated substrates (closed squares)). While baculovirus-expressed *Dnmt*1 polypeptide or 3T3 cell extract showed much higher activities when hemimethylated DNA was used as a substrate, Dnmt3a, Dnmt3b1 and Dnmt3b2 polypeptides showed no detectable preference for hemimethylated DNA (Fig. 8).

EXAMPLE 5

Two Dnmt3a Isoforms Produced from Alternative Promoters Show Different Subcellular Localization and Tissue Expression Patterns

Materials and Methods

[0255] Vectors: The GFP-Dnmt3, the Dnmt3-pcDNA, and the His_6 -tagged Dnmt3a constructs were generated by subcloning the corresponding Dnmt3a or Dnmt3b cDNA into pEGFP-C1 (Clontech), pcDNA6/V5-HisA (Invitrogen), and pET-28b(+) (Novagen), respectively. The P2 targeting vector was constructed by sequentially subcloning *Dnmt3a* genomic fragments, the hCMV-hygTK cassette, and the PGK-DTA cassette into pBluescript II SK. The *Dnmt3a* genomic fragments (left arm, 3.7 kb; right arm, 3.0 kb) were generated by PCR using a BAC clone (Genome Systems Inc.) as the template and the following pairs of oligonucleotides as primers: 5'-CTGGAATTCTCCTACCTTTG-3' (SEQ ID NO:95) and 5'-CCTGGATCCCAGCCAGTGAGCTGG-3' (SEQ ID NO:96) (for left arm), 5'-GTTCCGCGGCTGCTCATT-3' (SEQ ID NO:97) and 5'-CCACCGCGGCCGACTTGCCTCTACTTC-3' (SEQ ID NO:98) (for right arm). (The restriction sites used for cloning are underlined). The identities of the constructs were verified by DNA sequencing.

[0256] Antibodies: The Dnmt3 rabbit polyclonal antibodies, 164 and 157, were generated against mouse Dnmt3a amino acids 15-126 and Dnmt3b amino acids 1-181, respectively. The Dnmt3a mAb (clone 64B1446) was purchased from Imgenex. Anti-GFP mAb (a mixture of clones 7.1 and 13.1) was obtained from Roche. Anti-tubulin mAb (Ab-1) was obtained from Oncogene Research

Products. Anti-DNMT1 (human) polyclonal AB was purchased from New England Biolabs. Anti-histone H1 (AE-4) and anti-lamin B (M-20) were obtained from Santa Cruz Biotechnology.

[0257] Protein expression and analysis: Transient transfection was carried out in COS-7 or NIH 3T3 cells using LIPOFECTAMINE PLUS reagent (Invitrogen). Immunoprecipitation, immunoblotting, and fluorescence microscopy analyses were performed as previously described (He, D. *et al., J Cell Biol* 110, 569-580 (1990); Chen, T., and Richard, S. *Mol Cell Biol* 18 (8), 4863-71 (1998); Chen, T. *et al., Mol Biol Cell* 10 (9), 3015-33 (1999)).

[0258] Luciferase reporter assay: Luciferase reporter constructs as well as pGL-3-Basic (empty vector) were individually co-transfected with pRL-TK (internal control, Promega) into ES cells or NIH 3T3 cells. The cell lysates were analyzed for luciferase activities using the dual-luciferase reporter assay system (Promega).

[0259] 5' RACE, RT-PCR, and Northern hybridization: 5' RACE was carried out on total RNA prepared from ES cells using the 5' RACE system (Invitrogen)withDnmt3a-specific primers: 5'-AGCTGCTCGGCTCCG GCC-3' (SEQ ID NO:99) (for reverse transcription), 5'-TCCCCCACACCAGCTCTCC-3'(SEQ ID NO:100) (for 1st round PCR), and 5'-CTGCAATTACCTTGGCTT-3' (SEQ ID NO:101) (for 2nd round PCR). For RT-PCR analysis, total RNA was reverse transcribed with oligo(dT)$_{12-18}$ and the resulting cDNAs were amplified by PCR. Dnmt3a-specific primers used are 5'-TCCAGCGGCCCCGGGGAC-3' (SEQ ID NO:102) (F1), 5'-CCCAACCTGAGGAAGGGA-3' (SEQ ID NO:103)(F2), 5'-ACCAACATCGAATCCATG-3' (SEQ ID NO:104) (F3), 5'-TCCCGGGGCCGACTGCGA-3' (SEQ ID NO:105) (F4), 5'-AGGGGCTGCACCTGGCCTT-3' (SEQ ID NO:106) (F5), 5'-TCCCCCACACCAGCTCTCC-3' (SEQ ID NO:107) (R1), and 5'-CCTCTGCAGTACAGCTCA-3' (SEQ ID NO:108) (R2). Dnmt3b-specific primers used are 5'-TGGGATCGAGGGCCTCAAAC-3' (SEQ ID NO:109) and 5'-TTCCACAGGACAAACAGCGG-3'(SEQ ID NO:110) (for exon 10), 5'-GCGACAACCGTCCATTCTTC-3' (SEQ ID NO:111) and 5'-

CTCTGGGCACTGGCTCTGACC-3' (SEQ ID NO:112) (for exons 21 and 22). Northern hybridization was performed according to standard protocols. *Dnmt3a* cDNA fragments used as probes were generated by PCR. The primer pairs used were 5'- GCAGAGCCGCCTGAAGCC-3' (SEQ ID NO:113) and 5'-CCTTTTCCAACGTGCCAG-3'(SEQ ID NO:114) (for probe 1), and 5'-GCCAAGGTAATTGCAGTA-3' (SEQ ID NO:115) and 5'-GATGTTTCTGCACTTCTG-3' (SEQ ID NO:116) (for probe 2).

[0260] Targeted disruption of *Dnmt3a2* in ES cells. The P2 targeting vector was electroporated into *Dnmt3a*$^{+/-}$ ES cells (Okano, M. *et al., Cell* 99(3):247-257 (1999)), which were subsequently selected in hygromycin-containing medium. Genomic DNA isolated from hygromycin-resistant colonies was digested with *Sca*I and analyzed by Southern hybridization using a 0.45 kb *Kpn*I-*Spe*I fragment as a probe.

[0261] DNA methyltransferase assays. For in vitro DNA methyltansferase activity, His_6 -tagged Dnmt3a proteins were incubated with double-stranded poly(dI-dC) (Pharmacia) in the presence of S-adenosyl-L-methionine [methyl-^{3}H] (NEN), and the incorporation of ^{3}H methyl groups into poly(dI-dC) was measured as previously described (Okano, M. *et al., Nat. Genet. 19*(3):219-20 (1998)). For *de novo* methylation activity, human EC cell lines and breast/ovarian cancer cell lines were infected with Moloney murine leukemia virus, and the methylation status of newly integrated provirus was analyzed as previously described (Lei, H. *et al., Development 122*(10):3195-3205(1996)).

Results

Identification of Dnmt3b6 and Dnmt3a2

[0262] The Dnmt3a and Dnmt3b proteins show high sequence homology in the C-terminal catalytic domain, but they share little sequence similarity in the N-terminal regulatory region except for the conserved PWWP and PHD domains

(Fig. 11A). To characterize the Dnmt3 proteins, rabbit polyclonal antibodies were generated against the N-terminal regions of mouse Dnmt3a (antibody 164) and Dnmt3b (antibody 157), and a commercial monoclonal antibody (64B1446), which was raised against the full-length mouse Dnmt3a was also obtained. The epitope recognized by 64B1446 was mapped to a region (a.a. 705-908) at the C terminus. The specificity of the Dnmt3 antibodies was examined using GFP fusion proteins expressed in Cos-7 cells (Fig. 11B). Anti-GFP immunoblotting showed the expression of the GFP fusion proteins (1st panel). The polyclonal antibodies, 164 and 157, were specific for Dnmt3a and Dnmt3b, respectively (2nd and 3rd panels). The monoclonal antibody, 64B1446, reacted strongly with Dnmt3a proteins and weakly with Dnmt3b1 and Dnmt3b2, but not Dnmt3b3 (4th panel), consistent with the epitope-mapping results.

[0263] Previous studies showed that *Dnmt3a* and *Dnmt3b* transcripts were abundant in ES cells (Okano, M. *et al., Nat. Genet. 19*(3):219-220 (1998)), but their protein products had not been analyzed. To address this question, wild-type (J1), *Dnmt3a*$^{-/-}$ (6aa), *Dnmt3b*$^{-/-}$ (8bb), and *[Dnmt3a*$^{-/-}$, *Dnmt3b*$^{-/-}$] (7aabb) mutant ES cells (Okano, M. *et al., Cell 99*(3):247-257(1999)) were analyzed by immunoblotting with the Dnmt3 antibodies (Fig. 11C and 11D). Two distinct bands, which migrated at ~120 and ~110 kDa, were detected by antibody 157 in J1 and 6aa cells, but not in 8bb and 7aabb cells (Fig. 11C), indicating that these bands represent Dnmt3b proteins. The more abundant 120-kDa band most likely represents Dnmt3b1 and the 110-kDa band represents an isoform smaller than Dnmt3b2 but slightly larger than Dnmt3b3 (Fig. 11C). RT-PCR analysis confirmed the expression of two major *Dnmt3b* transcripts in ES cells; one corresponds to *Dnmt3b1* and the other is an alternatively spliced variant that lacks exons 21 and 22 (Fig. 16 and data not shown). This new isoform was named *Dnmt3b6* (schematically shown in Fig. 11A). Indeed, the 110-kDa band observed in ES cells co-migrated with protein expressed from Dnmt3b6 cDNA (Fig. 11C, lanes 8 and 9). Dnmt3b6 lacks motif IX and thus may not be enzymatically active, like Dnmt3b3 (Aoki, A. *et al., Nucleic Acids Res* 29 (17), 3506-12 (2001)).

[0264] Dnmt3a-specific antibody 164 detected a single band of ~130 kDa in J1 and 8bb cells, which co-migrated with the control Dnmt3a protein (Fig. 11D, lanes 1, 2 and 5), but not in 6aa and 7aabb cells (lanes 3 and 4). Surprisingly, when the same blot was reprobed with anti-Dnmt3a monoclonal antibody 64B1446, two more intense bands of ~120 kDa and ~100 kDa were detected in addition to the 130-kDa Dnmt3a protein in J1 cells (Fig. 1D, lane 7). The 120-kDa band represents Dnmt3b1 as it was also present in 6aa cells but absent in 8bb cells (lanes 9 and 10). Like the 130-kDa Dnmt3a protein, the 100-kDa band could be detected in 8bb cells (lane 10) but not in 6aa and 7aabb cells (lanes 8 and 9), indicating that it is a novel product of the *Dnmt3a* gene. We named this short form Dnmt3a2. Importantly, the immunoblotting result indicates that Dnmt3a2 is the predominant *Dnmt3a* gene product in ES cells (Fig. 11D).

[0265] The fact that Dnmt3a2 could not be recognized by antibody 164 suggests that Dnmt3a2 lacks the N-terminal region of Dnmt3a. Inspection of the *Dnmt3a* cDNA sequence revealed that, in addition to the known initiation codon (ATG1), two downstream in-frame ATGs (ATG2 and ATG3), corresponding to Met 159 and Met 220, were found to be within the Kozak consensus sequence. To test the possibility that Dnmt3a2 was produced by translation initiated at one of these ATGs, we expressed in 6aa cells two Dnmt3a proteins with the N-terminal 158 and 219 amino acids truncated and showed that Dnmt3a (220-908) co-migrated with endogenous Dnmt3a2 from J1 cells (Fig. 11E, compare lanes 3 and 4). This suggests that ATG3 might be the initiation codon for Dnmt3a2. To further determine whether Dnmt3a2 is produced from the same mRNA transcript as Dnmt3a, we transfected 6aa cells with an expression vector containing the entire *Dnmt3a* coding sequence. Immunoblotting analysis using antibody 64B1446 showed that only Dnmt3a was expressed (Fig. 11F, lane 2). These results suggest that Dnmt3a2 does not derive from *Dnmt3a* transcript by the use of an alternative ATG or from Dnmt3a protein by proteolytic cleavage or degradation.

Dnmt3a2 is encoded by transcripts initiated from a downstream promoter

[0266] To determine whether Dnmt3a and Dnmt3a2 are encoded by distinct mRNA transcripts, total RNA from J1, 6aa ES cells, and NIH 3T3 cells (which express only Dnmt3a, see Fig. 17) was analyzed by Northern hybridization with *Dnmt3a* cDNA probes upstream or downstream of ATG3 (Fig. 12B). The downstream probe (Probe 2, Fig. 12A) detected two major transcripts of 4.2 kb and 4.0 kb and a weak band of 9.5 kb from J1 cells (Fig. 12B, lane 5), consistent with our previous results (Okano, M. *et al., Nat Genet* 19 (3), 219-20 (1998)). All the transcripts were smaller and the intensity of 4.2 kb and 4.0 kb bands was substantially reduced in 6aa cells (lane 6), indicating that truncated transcripts were generated. The 9.5-kb transcript was also present at low level in NIH 3T3 cells, but the 4.2 kb and 4.0 kb transcripts were absent (lane 4). Interestingly, the upstream probe (Probe 1, Fig. 12A) recognized the 9.5 kb transcript in NIH 3T3 and J1 cells and a 7.5 kb truncated form in 6aa cells, but it failed to hybridize to the 4.2 kb and 4.0 kb transcripts in J1 cells (lanes 1-3). Taken together, these observations suggest that Dnmt3a2 is probably encoded by the 4.2 kb and 4.0 kb transcripts. Our previous data indicated that the 4.2 kb and 4.0 kb transcripts differ in their 3'UTR, probably due to alternative 3' processing (Okano, M. *et al., Nat. Genet. 19*(3):219-220 (1998)).

[0267] To determine the identity of the *Dnmt3a* transcripts, 5' RACE was performed on RNA prepared from J1 ES cells with primers annealing to *Dnmt3a* sequences downstream of the putative Dnmt3a2 translation start site (ATG3 at M220). Two species of *Dnmt3a* transcripts were obtained. One of them matched the *Dnmt3a* cDNA sequence and the other contained a 55-bp sequence at its 5' end that did not match any known *Dnmt3a* cDNA sequence. Searches of the Celera mouse genome database revealed that the 55-bp sequence was part of an exon located in an intron of the *Dnmt3a* gene. Using the new exon sequence as query, a mouse EST clone was identified, BE855330, which extended the exon to at least 117 bp. Sequencing analysis revealed that the EST clone shared all the

downstream exons with *Dnmt3a* (Fig. 12A). It is concluded that the newly identified transcript encodes Dnmt3a2 as its open reading frame would predict a protein that lacks the N-terminal 219 amino acids of Dnmt3a (Fig. 12A). As illustrated in Fig. 12A, the murine *Dnmt3a* gene consists of 24 exons. Exons 8-24 are shared by both *Dnmt3a* and *Dnmt3a2*. Exons 1-6 are present only in *Dnmt3a* whereas exon 7 (indicated by a *) is unique to *Dnmt3a2*.

[0268] The 5' RACE results were confirmed by RT-PCR analysis of total RNA from J1 cells using primers annealing to different *Dnmt3a* exons (Fig. 12A). Combination of *Dnmt3a*-specific (F1-F4) or *Dnmt3a2*-specific (F5) primers with a downstream primer in exon 9 (R1) verified the expression of both *Dnmt3a* and *Dnmt3a2* transcripts in ES cells (Fig. 12C, lanes 1-4 and 9-16). However, combination of the same *Dnmt3a* primers (F1-F4) with a primer in the unique *Dnmt3a2* exon (R2) failed to generate any PCR products (lanes 5-8). These results indicate that it is unlikely that the *Dnmt3a* and *Dnmt3a2* transcripts are produced via alternative splicing.

[0269] The nucleotide and predicted amino acid sequences of *Dnmt3a2* are presented in Figure 13A and B. By RT-PCR analysis and database searches, human *DNMT3A2* was also identified (Fig. 12A). The Nucleotide and predicted amino acid sequences of human *DNMT3A2* are presented in Figure 13C and D. An alignment of the human and murine cDNA sequences reveals strong similarity (Fig. 13E1-E4) except that human *DNMT3A2* contains an additional sequence of 68 bp in the 5'UTR, which is encoded by an extra exon located ~2.5 kb downstream of exon 7 (the newly identified exons are indicated by * in Fig. 12A). The predicted mouse Dnmt3a2 and human DNMT3A2 proteins, each consisting of 689 amino acids (Fig. 13B and D, respectively), show high sequence identity (Fig. 13F; 98.5%).

[0270] The observation that the *Dnmt3a2*-specific exon is located in a region >80 kb downstream of the putative *Dnmt3a* promoter suggests that *Dnmt3a2* transcription may be driven by a different promoter. Indeed, analysis of the large (~18 kb) "intron" preceding exon 7 with PROSCAN

(http://bimas.dcrt.nih.gov/molbio/proscan) predicted that a 1.4-kb region immediately upstream of exon 7 has high probability to function as a promoter. It should also be noted that the unique *Dnmt3a2* exon resides in a GC-rich CpG island, which is a hallmark of the promoter region of genes. The transcriptional activity of the putative promoter was tested using a reporter system (Fig. 14). A ~2.0 kb genomic fragment that includes the putative promoter (P2) was inserted, in both orientations, upstream of the cDNA encoding the firefly luciferase followed by the SV40 late poly(A) signal (Fig. 14A; *See* Fig. 27 for nucleotide sequence of the genomic fragment). Transient transfection experiments demonstrated that the P2 fragment has high promoter activity in ES cells but much lower activity in NIH 3T3 cells (Fig. 14B, P2-luc), consistent with the expression levels of *Dnmt3a2* in these cell types (Fig. 12B). The transcriptional activity of the P2 fragment is orientation-dependent, as the same fragment showed no promoter activity when subcloned in reverse orientation (Fig. 14B; P2R-luc). As a positive control, SV40 promoter worked equally well in both cell types. These data strongly suggest that the region 5' adjacent to exon 7 functions as a promoter and drives the expression of *Dnmt3a*2.

[0271] To confirm that exon 7 and the adjacent promoter are essential for the expression of *Dnmt3a*2, we deleted the P2 region from the wild-type allele in *Dnmt3a*$^{+/-}$ ES cells (Okano, M. *et al., Cell* 99 (3), 247-57 (1999)) by gene targeting. An hCMV-hygTK cassette was inserted in the opposite orientation of *Dnmt3a* transcription to avoid disruption of the *Dnmt3a* transcripts (Fig. 15A). We, therefore, expected that the removal of these sequences would abolish the transcription of *Dnmt3a*2, but not *Dnmt3*a. One clone (296) with deletion of the wild type allele was successfully isolated (Fig. 15B). As expected, Northern hybridization showed that the 4.2 kb and 4.0 kb transcripts were completely abolished in clone 296 cells (Fig. 15C). Consistently, immunoprecipitation and immunoblotting analyses demonstrated that Dnmt3a2 protein was abolished whereas Dnmt3a protein was produced in clone 296 cells at similar levels as in *Dnmt3a*$^{+/-}$ cells (Fig. 15D). These data provide genetic evidence that the newly

identified Dnmt3a2 is indeed encoded by mRNA transcribed from a downstream promoter.

Dnmt3a2 and Dnmt3a show similar methyltransferase activity but different subcellular localization patterns

[0272] To test whether Dnmt3a2 has methyltransferase enzyme activity, we generated recombinant Dnmt3a proteins and measured their catalytic activity by a standard in vitro methylation assay. Dnmt3a, Dnmt3a:PC→VD, and Dnmt3a2 were expressed in E. coli as N-terminally His_6-tagged fusion proteins and purified by metal chelation chromatography. The proteins were ~90% pure, as estimated by Coomassie blue staining (Fig. 16A, lanes 1-3), and their identity was verified by immunoblotting (lanes 4-6). As shown previously (Okano, M. *et al., Nat Genet* 19 (3), 219-20 (1998)) Dnmt3a was able to transfer methyl groups to double-stranded poly (dI-dC). Mutation of the PC motif in the catalytic domain (Dnmt3a:PC→VD) abolished the activity. Dnmt3a2 showed similar enzyme activity as Dnmt3a (Fig. 16B), demonstrating that Dnmt3a2 is an active DNA methyltransferase.

[0273] It has been recently reported that Dnmt3a localizes to heterochromatin (Bachman, K. E. *et al., J Biol Chem* 276 (34), 32282-7 (2001)). To determine whether Dnmt3a2 localizes differently from Dnmt3a, GFP-Dnmt3a fusion proteins were expressed in NIH 3T3 cells and their localization was analyzed by fluorescence microscopy. Dnmt3a localized exclusively in the nuclei and concentrated in nuclear foci that correspond to DAPI (4,6-diamidino-2-phenylindole) bright spots, consistent with heterochromatin association. In contrast, Dnmt3a2 showed a diffused pattern excluding nucleoli and heterochromatin. Although Dnmt3a2 localized mainly in the nuclei, weak staining was also observed in the cytoplasm (Fig. 16C). Similar results were obtained when the GFP fusion proteins were expressed in ES cells. These data indicate that the N-terminal 219 amino acids of Dnmt3a are required for its exclusive nuclear localization and heterochromatin association.

[0274] To confirm the localization data, we investigated the subcellular distribution of endogenous Dnmt3 proteins. ES cells were extracted sequentially to obtain the cytoplasmic, chromatin, and nuclear matrix fractions. Immunoblotting analysis with antibody 64B1446 showed that Dnmt3a and Dnmt3a2 as well as Dnmt3b1 fractionate mainly with chromatin and small proportions of these proteins also associate with the nuclear matrix (Fig. 16D). While Dnmt3a and Dnmt3b1 were exclusively nuclear, a significant proportion of Dnmt3a2 was present in the cytoplasmic fraction (Fig. 16D), consistent with the localization results (Fig. 16C). The efficacy of the fractionation procedure was verified by immunoblotting with control antibodies specific to histone H1 (a component of chromatin) and lamin B (a nuclear matrix-associated protein) (Fig. 16D). Taken together, these results suggest that Dnmt3a associates mainly with heterochromatin and Dnmt3a2 associates primarily with euchromatin.

Expression of Dnmt3a2 and Dnmt3b in mouse tissues and human cell lines correlate with *de novo* methylation activity

[0275] Since *de novo* methylation activity changes during differentiation, the levels of Dnmt3a and Dnmt3b proteins in differentiating ES cells were examined. ES cells were differentiated as embryoid bodies in vitro for 14 days and the change of Dnmt3 protein levels was monitored by immunoblotting (Fig. 17A). Dnmt3a, Dnmt3a2, and Dnmt3b were all upregulated upon differentiation, with the highest level observed in embryoid bodies at 4-6 days. However, after 6 days of differentiation, the level of Dnmt3a2 and Dnmt3b rapidly decreased, whereas the level of Dnmt3a sustained throughout the course of the experiment.

[0276] The expression of Dnmt3a and Dnmt3b proteins in somatic tissues from 3-week-old mice was then examined by immunoprecipitation and immunoblot analysis. As shown in Fig. 17B, Dnmt3a was detected in all tissues except for small intestines, whereas Dnmt3a2 and Dnmt3b expression was more restricted. Both Dnmt3a2 and Dnmt3b proteins were detected in testis, spleen, and thymus, tissues known to contain cells that undergo active *de novo* methylation. Dnmt3b

was also present at low levels in liver (Fig. 17B). RT-PCR analysis confirmed the immunoblotting results and also revealed the expression of *Dnmt3a2* and *Dnmt3b* in ovary (Fig. 17C and 17D). Based on the presence or absence of *Dnmt3b* exon 10 and/or exons 21/22, we were able to determine the *Dnmt3b* isoforms (Fig. 17D). Therefore, the Dnmt3b doublets observed in testis, spleen, thymus, and liver (Fig. 17B) most likely represent Dnmt3b2 and Dnmt3b3. Of note is that the relative levels of Dnmt3b2 and Dnmt3b3 are different in these tissues (Fig. 17B). Although Dnmt3b proteins could not be detected in many tissues (Fig. 17B), low levels of *Dnmt3b* transcripts (mainly *Dnmt3*b3) were expressed ubiquitously (Fig. 17D). *Dnmt3b1* and.16 *Dnmt3b6* were detected only in ES cells (Fig. 17D). These observations, along with the dynamic changes during ES cell differentiation, indicate that *Dnmt3a2* and *Dnmt3b* are coordinately regulated and their expression correlates with *de novo* methylation activity.

[0277] Since overexpression of *DNMT*1, *DNMT3*A, and *DNMT3B* transcripts have been reported in various human cancers, the expression of various DNMT proteins was examined in embryonal carcinoma and breast/ovarian cancer cell lines by immunoblotting. We showed that five EC cell lines expressed relatively high levels of DNMT3A2 and low levels of DNMT3A (Fig. 18A). DNMT3B was also highly expressed in these cells but different cells expressed different isoforms (Fig. 18B). In several breast and ovarian cancer cell lines, DNMT1 was expressed at comparable levels, which was similar to the level in an EC cell line, NCCIT (Fig. 18C, 1st panel) (note that the antibody does not recognize mouse Dnmt1 in J1 and NIH 3T3 cells). Low levels of DNMT3A1 were detected in most cell lines (Fig. 18C, 2nd panel). Although DNMT3A2 and DNMT3B proteins were also detectable in most of the breast/ovarian cancer cell lines, their levels were very low as compared to EC and ES cells (Fig. 18C, 3rd and 4th panels).

[0278] It was then investigated whether the expression levels of DNMT proteins correlate with *de novo* methylation activity. Human EC and breast and ovarian cancer cell lines were infected with Moloney murine leukemia virus (MMLV, Fig. 18D, lower panel), and the methylation status of proviral DNA was analyzed

using the CpG methylation sensitive enzyme *Hpa* II (Fig. 18D). The proviral DNA was partially or completely methylated in the EC cell lines, as indicated by the presence of *Hpa II*-resistant bands ranging from 0.8 kb (unmethylated band) to 1.3 kb (fully methylated band), and the level of methylation increased with time (lanes 1-13, compare day 5 and day 20). In contrast, little or no *de novo* methylation activity was detected in any of the breast and ovarian cancer cell lines examined (lanes 14-21). Since DNMT1 was readily detected in all the cell lines (Fig. 17C), the results provide additional evidence that DNMT1 does not have *de novo* methyltransferase activity, consistent with the current view that it functions as a maintenance enzyme. It is also unlikely that DNMT3A1 caused the difference in *de novo* methylation between EC cell lines and breast/ovarian cancer cells, as the expression level of DNMT3A is low but similar in both groups of cell lines (Fig. 18C). The absence of DNMT3B1/3B2 in several EC cell lines (PA-1, NTERA-2, and Tera-2) suggested that the *de novo* methylation activity observed in these cells can be attributed to the activity of DNMT3A2. The results are therefore most consistent with the notion that DNMT3A2 and DNMT3B1/3B2 are responsible for active *de novo* methylation of provirus DNA in ES cells and EC cells.

Discussion

[0279] In this study it was demonstrated that the *Dnmt3a* gene encodes at least two isoforms, termed Dnmt3a and Dnmt3a2, of approximately 130 kDa and 100 kDa, respectively. The newly identified Dnmt3a2 protein, which lacks the N-terminal region of Dnmt3a, is encoded by transcripts initiated from a downstream promoter and represents the major isoform in ES cells and EC cells. This conclusion is supported by several lines of evidence from molecular and genetic analyses of wild type and Dnmt3a-deficient ES cells. First, antibodies specific to the N-terminal region of Dnmt3a failed to detect the 100-kDa protein in ES cells and a 5' cDNA probe upstream of the first coding exon of *Dnmt3a2* failed to

hybridize to the major 4.0 kb and 4.2 kb transcripts. Second, 5' RACE and RT-PCR analysis identified a 5' exon upstream of the *Dnmt3a2* coding region, which is located in a large intron of *Dnmt3*a. Third, a GC-rich "intronic" region upstream of the *Dnmt3a2*-specific exon showed strong promoter activity for the expression of a reporter gene in ES cells and much lower activity in NIH 3T3 cells, consistent with *Dnmt3a2* expression status in these cells. Finally, deletion of the putative promoter region abolished *Dnmt3a2* transcripts and Dnmt3b2 protein, whereas transcription and translation of *Dnmt3a* were unaffected.

[0280] While both Dnmt3a and Dnmt3a2 are active DNA methyltransferases as shown by in vitro assays, they differ from one another in two main features. First, Dnmt3a2 showed a diffused nuclear staining pattern excluding heterochromatin, in contrast to Dnmt3a, which is concentrated in heterochromatin. It is believed that Dnmt3a and Dnmt3a2 may modify different chromatin domains, with Dnmt3a preferentially methylating heterochromatin and Dnmt3a2 preferentially methylating euchromatin. Given that hypermethylation of single-copy genes, which usually reside in euchromatic regions, contributes to diseases such as cancers, the association of Dnmt3a2 with euchromatin may potentially link Dnmt3a2 action to ontogenesis. Notably, Dnmt3a2 is detectable in many breast/ovarian cancer cell lines although the expression level is not sufficient to cause *de novo* methylation of provirus (Fig. 18). Second, expression of Dnmt3a2 is developmentally regulated, whereas Dnmt3a is ubiquitously expressed. It was observed that Dnmt3a2 is expressed only in tissues, such as testis, ovary, spleen, and thymus, in which *de novo* methylation is believed to occur during cellular differentiation. Analysis of *de novo* methylation activity in human cell lines also suggested that DNMT3A2 is capable of methylating newly integrated retroviral DNA. Therefore, Dnmt3a2 may function as a *de novo* methyltransferase. The absence of Dnmt3a2 in most somatic tissues suggests that expression of Dnmt3a2 must be tightly regulated to avoid abnormal *de novo* methylation, which could be toxic to cells. Consistent with these results, it was observed that it was difficult

to establish stable cell lines with overexpression of Dnmt3a2, but not when Dnmt3a or mutated Dnmt3a2 (mutation of the PC motif) was overexpressed.

[0281] In this study, a novel isoform of Dnmt3b, termed Dnmt3b6 was also identified. It was demonstrated that different Dnmt3b isoforms exhibit different tissue distributions. Dnmt3b1 and Dnmt3b6 are the predominant forms in ES cells, while Dnmt3b2 and Dnmt3b3 are expressed at relatively high levels in testis, ovary, spleen, thymus, and liver. It is believed that Dnmt3b1 and Dnmt3b2 function as *de novo* methyltransferases, whereas Dnmt3b3 and Dnmt3b6 function as regulators of DNA methylation.

[0282] Genetic studies have shown that *Dnmt3a* and *Dnmt3b* are essential for *de novo* methylation in ES cells and during embryonic development (Okano, M. *et al., Cell* 99 (3), 247-57 (1999)). Since Dnmt3a and Dnmt3b isoforms show different biochemical properties and expression patterns, they may have distinct functions in development. Dnmt3a2 and Dnmt3b1 are the major isoforms detected in ES cells and likely have redundant functions in carrying out *de novo* methylation of provirus DNA (Okano, M. *et al., Cell* 99 (3), 247-57 (1999)). Interestingly, the expression level of both Dnmt3a and Dnmt3a2, and different Dnmt3b isoforms is elevated during early stages of ES cell differentiation, but only Dnmt3a expression persists to the late differentiation stage, reminiscent of Dnmt3a and Dnmt3b expression in embryos (Okano, M. *et al., Cell* 99 (3), 247-57 (1999)). It is believed that Dnmt3a2 and Dnmt3b1/3b2 may be involved in *de novo* methylation in early postimplantation embryos. While these enzymes may have overlapping functions in modifying various genomic sequences, protein targeting may confer specificity to them as well. Lack of access to heterochromatin may explain why Dnmt3a2 can not compensate for Dnmt3b in methylating centromeric minor satellite repeats (Okano, M. *et al., Cell* 99 (3), 247-57 (1999)). Dnmt3a2 and Dnmt3b are also expressed at relatively high levels in testis, ovary, spleen and thymus and may play an important role in regulation of genomic imprinting, gametogenesis, and lymphocyte differentiation. It has been shown that disruption of both *Dnmt3a* and *Dnmt3a2* by deleting the

conserved motifs in the catalytic domain perturbs *de novo* methylation of maternally imprinted genes during oocyte maturation and spermatogenesis (Hata, K. *et al., Development* 129, 1983-93). Dnmt3a (and Dnmt3b3) is expressed at low levels in most tissues and cell lines analyzed, suggestive of a housekeeping function.

EXAMPLE 6

Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b

[0283] DNA methyltransferases Dnmt3a and Dnmt3b carry out de novo methylation of the mouse genome during early postimplantation development and of maternally imprinted genes in the oocyte. In this study, it is shown that Dnmt3a and Dnmt3b are also essential for the stable inheritance, or 'maintenance' of DNA methylation patterns. Inactivation of both Dnmt3a and Dnmt3b in ES cells results in progressive loss of methylation in various repeats and single copy genes. Interestingly, introduction of various Dnmt3a and Dnmt3b isoforms back into highly demethylated mutant ES cells restores genomic methylation patterns and different isoforms have both common and specific DNA targets, but they all fail to restore the maternal methylation imprints. Evidence is provided shows that Dnmt3b3 (and 3b6 as well) has no enzymatic activity in vivo, but may function as a negative regulator of DNA methylation. It is also shown that hypermethylation of genomic DNA by Dnmt3a and Dnmt3b is necessary for ES cells to form teratomas in nude mice. These results indicate that genomic methylation patterns are determined partly through differential expression of different Dnmt3a and Dnmt3b isoforms.

Introduction

[0284] DNA methylation is essential for mammalian development and plays crucial roles in a variety of biological processes such as genomic imprinting and X chromosome inactivation (Li, E. *Nat Rev Genet* 3:662-73 (2002)). DNA methylation patterns are established during embryonic development through a highly orchestrated process that involves demethylation and de novo methylation and can be inherited in a clonal fashion through the action of maintenance methyltransferase activity (Bird, A. P., and A. P. Wolffe. *Cell* 99:451-4 (1999); Li, E. *Nat Rev Genet* 3:662-73 (2002); Reik *et al.*, *Science* 293:1089-93 (2001)). During preimplantation development, both the paternal and maternal genomes undergo a wave of demethylation, which erases most of the methylation patterns inherited from the gametes. Shortly after implantation, the embryo undergoes a wave of de novo methylation, which establishes a new methylation pattern (Howlett, S. K., and W. Reik. *Development* 113:119-27 (1991); Kafri *et al.*, *Genes Dev* 6:705-14 (1992); Monk *et al.*, *Development* 99:371-82 (1987); Sanford *et al.*, *Genes Dev* 1:1039-46 (1987)). De novo methylation also occurs during gametogenesis in both male and female germ cells and is believed to play a critical role in the establishment of genomic imprinting in the gametes. Genomic imprinting is an epigenetic process that marks alleles according to their parental origin during gametogenesis and results in monoallelic expression of a small set of genes, known as imprinted genes, in the offspring (Jaenisch, R. *Trends Genet* 13:323-9 (1997); Li, E. *Nat Rev Genet* 3:662-73 (2002); Reik, W., and J. Walter. *Nat Rev Genet* 2:21-32 (2001)). De novo methylation activity is present mainly in embryonic stem (ES) cells and embryonal carcinoma (EC) cells, early postimplantation embryos, and developing germ cells, whereas it is largely suppressed in differentiated somatic cells (Kafri *et al.*, *Genes Dev* 6:705-14 (1992); Lei *et al.*, *Development* 122:3195-205 (1996); Santos *et al.*, *Dev Biol* 241:172-82 (2002); Stewart *et al.*, *Proc Natl Acad Sci USA* 79:4098-102 (1982)).

Therefore, ES cells can be a good model system for studying the mechanisms of de novo methylation.

[0285] Three active DNA cytosine methyltransferases, namely Dnmt1, Dnmt3a, and Dnmt3b, have been identified in human and mouse (Bestor *et al.*, *J Mol Biol* 203:971-83 (1988); Okano *et al.*, *Nat Genet* 19:219-20 (1998); Xie *et al.*, *Gene* 236:87-95 (1999)). Dnmt1 is ubiquitously expressed in proliferating cells and localizes to DNA replication foci (Leonhardt *et al.*, *Cell* 71:865-73 (1992)). Purified Dnmt1 protein methylates hemi-methylated DNA substrates more efficiently than unmethylated DNA in vitro (Bestor, T. H. *EMBO J* 11:2611-7 (1992)). Despite its activity in vitro, Dnmt1 has not been convincingly shown to be able to initiate de novo methylation in vivo. Moreover, inactivation of Dnmt1 in ES cells and mice leads to extensive demethylation of all sequences examined (Lei *et al.*, *Development* 122:3195-205 (1996); Li *et al.*, *Cell* 69:915-26 (1992)). All these findings suggest that Dnmt1 functions primarily as a maintenance methyltransferase that is responsible for copying the parental-strand methylation pattern onto the daughter strand after each round of DNA replication. In contrast, Dnmt3a and Dnmt3b are highly expressed in ES cells, early embryos, and developing germ cells, but expressed at low levels in differentiated somatic cells (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002); Okano *et al.*, *Nat Genet* 19:219-20 (1998)). Indeed, genetic studies have demonstrated that Dnmt3a and Dnmt3b are essential for de novo methylation in ES cells and postimplantation embryos as well as for de novo methylation of imprinted genes in the germ cells (Hata *et al.*, Development 129:1983-93 (2002); Okano *et al.*, *Cell* 99:247-57 (1999)). Although Dnmt3a and Dnmt3b function primarily as de novo methyltransferases to establish methylation patterns, they may also play a role in maintaining methylation patterns. We have previously shown that some genomic sequences, such as the differentially methylated region 2 (*DMR2*) of *Igf2* and the 5' region of *Xist*, are almost completely demethylated and an L1-like repeat is partially demethylated in mutant ES cells that lack Dnmt3a and Dnmt3b (Liang *et al.*, *Mol Cell Biol* 22:480-91 (2002); Okano *et al.*, *Cell* 99:247-57 (1999)).

[0286] At least two Dnmt3a and six Dnmt3b isoforms have been identified (Fig. 20A) (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002); Hansen *et al.*, *Proc Natl Acad Sci USA* 96:14412-7 (1999); Okano *et al.*, *Nat Genet* 19:219-20 (1998); Robertson *et al.*, *Nucleic Acids Res* 27:2291-8 (1999); Xie *et al.*, *Gene* 236:87-95 (1999)). Dnmt3a and Dnmt3a2 are encoded by transcripts initiated from two different promoters. Dnmt3a2 lacks the N-terminal region of the full-length Dnmt3a and, as a result, they exhibit different subcellular localization patterns. While Dnmt3a is concentrated in heterochromatic foci, Dnmt3a2 localizes diffusely in the nucleus (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002)). Unlike the Dnmt3a isoforms, all the known Dnmt3b isoforms are derived from alternative splicing. Dnmt3b1 and Dnmt3b2 are enzymatically active, as shown by in vitro methyltransferase assays, whereas Dnmt3b3, which lacks part of motif IX, appears to be inactive (Aoki *et al.*, *Nucleic Acids* Res 29:3506-12 (2001); Okano *et al.*, *Nat Genet* 19:219-20 (1998)). Dnmt3b4, Dnmt3b5, and Dnmt3b6 are also presumably inactive because they lack either part of motif IX (Dnmt3b6) or both motifs IX and X (Dnmt3b4 and Dnmt3b5) (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002); Hansen *et al.*, *Proc Natl Acad Sci USA* 96:14412-7 (1999); Robertson *et al.*, *Nucleic Acids Res* 27:2291-8 (1999)). Like Dnmt3a, Dnmt3b1 has been shown to localize to heterochromatin (Bachman *et al.*, *J Biol Chem* 276:32282-7 (2001)). These Dnmt3a/3b isoforms show different expression patterns during development. Dnmt3a2 and Dnmt3b1 are highly expressed in ES cells and germ cells but almost undetectable in most somatic tissues, whereas Dnmt3a and Dnmt3b3 are expressed at low levels in almost all somatic tissues and cell lines examined (Beaulieu *et al.*, *J Biol Chem* 277:28176-81 (2001)).

[0287] In this study, we introduced various Dnmt3a/3b isoforms individually back into [Dnmt3a-/-, Dnmt3b-/-] mutant ES cells and showed that these isoforms have both shared and specific genomic targets. In addition, we demonstrated that Dnmt3a and Dnmt3b are required for stable inheritance of global DNA methylation patterns in ES cells and that maintenance of genomic methylation

above a threshold level, but not the presence of Dnmt3a and Dnmt3b proteins, is essential for ES cell differentiation and teratoma formation.

Materials and Methods

[0288] ES cell culture: Wild-type J1 and mutant ES cells were maintained in Dulbecco's modified Eagle medium (DMEM, Invitrogen) supplemented with 15% fetal bovine serum (HyClone), 0.1 mM non-essential amino acids (Invitrogen), 0.1 mM b-mercaptoethanol, 50 U/ml penicillin, 50 mg/ml streptomycin, and 500 U/ml leukemia inhibitory factor (LIF, Invitrogen). The cells were normally grown on gelatin-coated Petri dishes without feeder cells. For long-term culture, the cells were trypsinized and passaged every other day and the passage numbers were recorded.

[0289] DNA constructions: The plasmid vectors expressing Dnmt1, Dnmt3a, Dnmt3a2, Dnmt3b1, Dnmt3b3, and Dnmt3b1:PC (a mutant Dnmt3b1 with the proline-cysteine di-peptide at the active site substituted with glycine-threonine) were generated by subcloning the corresponding cDNAs into pCAG-IRESblast, an expression vector that contains a CAG promoter (a synthetic promoter that includes the chicken b-actin promoter and the human cytomegalovirus immediate early enhancer). pCAG-IRESblast was constructed by replacing the EcoRI-Xho I fragment of pCAGN2-R(H1)-S3H-I-ZF3 (gift from R. Jaenisch) with an IRES-blasticidin cassette.

[0290] The *Dnmt3b1* targeting vector, in which a 2-kb region containing exons 21 and 22 was replaced by the PGK-puromycin cassette, was generated by sequentially subcloning *Dnmt3b* genomic fragments (the 8-kb 5' arm and 3.3-kb 3'arm were both obtained from a BAC clone), the PGK-puromycin cassette, and the PGK-DTA cassette into pBluescript II SK. The identities of all constructs were verified by DNA sequencing.

[0291] Stable expression of DNA methyltransferases in ES cells: Expression vectors encoding Dnmt3a and Dnmt3b isoforms or Dnmt1 were electroporated

into [Dnmt3a-/-. Dnmt3b-/-] or Dnmt1-/- ES cells (Lei *et al.*, *Development* 122:3195-205 (1996); Okano *et al.*, *Cell* 99:247-57 (1999)), which were subsequently selected in blasticidin-containing medium for seven days. Blasticidin-resistant colonies were examined for protein expression by immunoblotting analysis using the following antibodies: monoclonal anti-Dnmt3a (clone 64B1446, Imgenex) (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002)), polyclonal anti-Dnmt3b (Chen *et al.*, *J Biol Chem* 277:38746-54 (2002)), or polyclonal anti-Dnmt1 (gift from S. Tajima). As loading controls, the levels of a-tubulin in these samples were determined by immunoblotting with monoclonal anti-tubulin antibody (Ab-1, Oncogene Research Products). Expression of the intended Dnmt proteins was observed in ~90% of the colonies, most of which maintained the expression level after four weeks of culture in blasticidin-containing medium.

[0292] Targeted disruption of Dnmt3b1 in ES cells: The Dnmt3b1 targeting vector was transfected into Dnmt3b+/- or [Dnmt3a-/-, Dnmt3b+/-] ES cells (Okano, M., *et al.*, *Cell 99*:247-257 (1999)) via electroporation and transfected cells were selected with puromycin. Genomic DNA isolated from puromycin-resistant colonies was digested with EcoRV and analyzed by Southern hybridization using a probe 3' external to the targeting construct. The targeting frequency for the wild-type allele in Dnmt3b+/- and [Dnmt3a-/-, Dnmt3b+/-] cells was 4/150 and 6/200, respectively.

[0293] DNA methylation analysis: Genomic DNA isolated from various ES cell lines was digested with methylation-sensitive restriction enzymes, and analyzed by Southern hybridization as previously described (Lei, H. *et al.*, *Development* 122:3195-3205 (1996)). Probes used for methylation analysis include the following: pMO for endogenous C-type retroviruses (Genbank accession NC_001501)(Li, E. *et al.*, *Cell* 69:915-926 (1992)), pMR150 for minor satellite repeats (accession X14469 X07949)(Chapman *et al.*, *Nature 307*:284-286 (1984)), IAP (accession AF303453)(Walsh *et al.*, *Nat Genet* 20:116-117 (1998)),

3' region of *β-globin* cDNA (accession J00413 K01748 K03545)(PCR product) (Dennis *et al.*, *Genes Dev* 15:2940-4 (2001)), 5' region of *Pgk-1* cDNA (accession M18735)(PCR product) (Dennis *et al.*, Genes Dev 15:2940-4 (2001)), coding region of *Pgk-2* cDNA (PCR product) (Dennis *et al.*, Genes Dev 15:2940-4 (2001)), 5' region of *Xist* cDNA (accession AJ421479, gift from T. Sado), the *H19* upstream region (accession U19619)(Tremblay *et al.*, *Nat Genet* 9:407-13 (1995)), *DMR2* or "probe 6" for *Igf2* (accession NM_010514)(Feil *et al.*, *Development* 120:2933-43 1994)), the *Igf2r* region 2 probe (accession NM_010515) (Stoger *et al.*, *Cell* 73:61-71 (1993)), *Peg1* (accession NM_008590)(Lefebvre *et al.*, *Hum Mol Genet* 6:1907-15 (1997)), *Snrpn DMR1* (accession NM_013670)(Shemer *et al.*, *Proc Natl Acad Sci USA* 94:10267-72 (1997)), and an oligonucleotide probe (5'-TAT GGC GAG GAA AAC TGA AAA AGG TGG AAA ATT TAG AAA TGT CCA CTG TAG GAC GTG GAA TAT GGC AAG-3' SEQ ID NO:117) specific to major satellite repeats.

Results

[0294] Inactivation of Dnmt3a and Dnmt3b results in progressive loss of DNA methylation in ES cells. Genetic studies have demonstrated that Dnmt3a and Dnmt3b carry out de novo methylation of the mouse genome during early embryonic development (Okano, M. *et al.*, *Cell* 99:247-257 (1999)). To investigate whether these enzymes are also involved in maintaining global DNA methylation patterns, we cultured [Dnmt3a-/-, Dnmt3b-/-] ES cells (Okano, M. *et al.*, *Cell* 99:247-257 (1999)) continuously for various periods of time and examined the methylation status of various genomic sequences using methylation-sensitive restriction enzymes. The endogenous C-type retroviruses and intracisternal A particle (IAP) repeats, which are interspersed in the mouse genome with about 100 and 1000 copies per haploid genome, respectively, are normally highly methylated in ES cells (Li, E. *et al.*, *Cell* 69:915-926 (1992); Okano, M. *et al.*, *Cell* 99:247-257 (1999)). These sequences became

progressively demethylated in two independent [Dnmt3a-/-, Dnmt3b-/-] cell lines (7aabb and 10aabb), as indicated by increasing sensitivity to Hpa II digestion (Fig. 19A). Similar results were obtained when DNA methylation of the major and minor satellite repeats was analyzed (Fig. 19A). The major and minor satellite repeats are located in the pericentromeric and centromeric regions at copy numbers of 700,000 and 50,000-100,000, respectively. After prolonged culture of [Dnmt3a-/-, Dnmt3b-/-] ES cells for about 5 months, DNA methylation in both repeats and unique genes examined was almost completely depleted (see below). No significant change in global methylation was observed when wild-type (J1) and Dnmt3a-/- (6aa) or Dnmt3b-/- (8bb) single mutant ES cells were grown in culture for the same periods of time (Fig. 19B, also see below). Loss of methylation in [Dnmt3a-/-, Dnmt3b-/-] ES cells was not due to reduced expression of Dnmt1 as immunoblotting analysis indicated that early-passage and late-passage cells had similar levels of Dnmt1 protein (Fig. 19C). These results suggested that the Dnmt3 family of methyltransferases are required for stable inheritance of global DNA methylation patterns in ES cells and Dnmt3a and Dnmt3b have largely redundant functions in this respect.

Stable expression of Dnmt3a and Dnmt3b in [Dnmt3a-/-, Dnmt3b-/-] ES cells restores DNA methylation

[0295] Dnmt3a and Dnmt3b isoforms show distinct expression profiles and cellular localization patterns (Bachman, K. E. *et al., J Biol Chem* 276:32282-32287 (2001); Chen, T. *et al., J Biol Chem* 277:38746-54 (2002)), raising the possibility that they may methylate different sets of sequences in the genome. To investigate whether the demethylated state of the [Dnmt3a-/-, Dnmt3b-/-] ES cell genome is reversible and whether different Dnmt3a and Dnmt3b isoforms have distinct specificities in re-establishing methylation patterns, we introduced cDNAs encoding Dnmt3a, Dnmt3a2, Dnmt3b1, Dnmt3b3, and Dnmt3b1:PC (Dnmt3b1 with its PC motif mutated) into late-passage 7aabb ES cells (Okano, M. *et al., Cell* 99:247-257 (1999)). DNA methyltransferases Dnmt3a and

Dnmt3b are essential for *de novo* methylation and mammalian development (Okano, M. *et al., Cell* 99:247-257 (1999)) by random integration. Each cDNA was subcloned in a plasmid vector in which a CAG promoter drives the expression of a bicistronic transcript that encodes both the intended Dnmt protein and the selection marker, blasticidin S deaminase (Fig. 20B, top panel). After selection with blasticidin, we were able to obtain individual clones that express various levels of Dnmt3a or Dnmt3b proteins, as determined by immunoblotting analysis (Fig. 20B). The monoclonal Dnmt3a antibody, which recognizes the C-terminal region of Dnmt3a (Fig. 20A), strongly reacts with Dnmt3a and Dnmt3a2 and weakly reacts with Dnmt3b1 and Dnmt3b2, but not the other Dnmt3b isoforms Chen, T., *et al.,* (Chen, T. *et al., J Biol Chem* 277:38746-54 (2002)). The polyclonal Dnmt3b antibody, which was raised against the N-terminal region of Dnmt3b (Fig. 20A), is Dnmt3b-specific and recognizes all known Dnmt3b isoforms (Chen, T. *et al., J Biol Chem* 277:38746-54 (2002)). For each construct, we chose two independent clones for methylation analysis. The relative levels of Dnmt3a/3b proteins expressed in these clones, as compared to the levels of the corresponding endogenous Dnmt3a/3b isoforms in wild-type ES cells (J1, 100%), were roughly estimated based on the intensity of the bands: Dnmt3a (clone 1: 500%, clone 2: 200%), Dnmt3a2 (clone 1: 150%, clone 2: 200%), Dnmt3b1 (clone 1: 150%, clone 2: 80%), Dnmt3b3 (clone 1: 400%, clone 2: 500%, compared with endogenous Dnmt3b6), and Dnmt3b1:PC (clone 1: 80%, clone 2: 50%, compared with endogenous Dnmt3b1)(Fig. 20B). We also confirmed by immunoblotting analysis that there was no cross-contamination between the control ES cell lines (J1, 6aa, 8bb, and 7aabb) during the course of long-term passage (Fig. 20B, middle and bottom panels, lanes 1-4).

[0296] We first examined whether repetitive elements could be re-methylated by the expressed Dnmt3a/3b proteins in 7aabb cells. As shown in Fig. 21A-D, expression of Dnmt3a, Dnmt3a2, or Dnmt3b1 substantially restored the methylation levels of the endogenous C-type retroviral DNA, the IAP repeats, and the major and minor satellite repeats, whereas expression of Dnmt3b3 or

Dnmt3b1:PC had no effect. While the two Dnmt3a isoforms showed similar efficiency in methylating these repetitive sequences, Dnmt3a/3a2 and Dnmt3b1 exhibited distinct sequence preferences. As compared to Dnmt3a/3a2, Dnmt3b1 was substantially more efficient in methylating the minor satellite repeats and slightly less efficient in methylating the major satellite repeats and the endogenous C-type retroviral DNA. These enzymes were equally efficient in methylating the IAP repeats and restored the methylation level to normal. To confirm these results, we analyzed genomic DNA from late-passage 6aa and 8bb ES cells and showed that the methylation patterns in these sequences were consistent with those observed in the corresponding Dnmt3a/3b stable clones.

[0297] To determine whether expression of Dnmt3a/3b proteins in 7aabb cells also affects methylation of unique genes, a number of specific genomic loci were examined. The *b-globin* and phosphoglycerate kinase 2 (*Pgk-2*) genes are highly methylated autosomal genes that show tissue-specific expression patterns. *Pgk-1* and *Xist*, two other highly methylated genes, are located on the X chromosome. The methylation-sensitive sites examined were located in the 5' region (*Pgk-1* and *Xist*), the coding region (*Pgk-2*), or the 3' region (*b-globin*) of the genes. All four loci were highly methylated in the wild type ES cells (J1) and became substantially demethylated in late-passage 7aabb cells (Fig. 21E-H). With expression of Dnmt3a, Dnmt3a2, or Dnmt3b1, but not Dnmt3b3 or Dnmt3b1:PC, in 7aabb cells, the examined regions in *b-globin*, *Pgk-1*, and *Pgk-2* genes were completely or partially re-methylated. These results were in agreement with the fact that methylation of these loci was maintained in 8bb and 6aa cells (Fig. 21E-G). Interestingly, Dnmt3a or Dnmt3a2 was able to restore methylation of the *Xist* promoter region to normal, but Dnmt3b1 was not (Fig. 21H). Consistently, inactivation of Dnmt3a alone in ES cells (6aa) resulted in demethylation of the *Xist* promoter region, whereas inactivation of Dnmt3b alone (8bb) had no effect (Fig. 21H), suggesting that Dnmt3a, but not Dnmt3b, is capable of establishing and is required for maintaining methylation of this particular region. Taken together, these data demonstrate that methylation of the highly demethylated

genome of [Dnmt3a-/-, Dnmt3b-/-] ES cells can be largely re-established by Dnmt3a and Dnmt3b and these enzymes have both shared and specific DNA targets.

Methylation of imprinted genes

[0298] Methylation of some imprinted genes, such as *H19* and Igf2 receptor (*Igf2r*), is maintained in early-passage [Dnmt3a-/-, Dnmt3b-/-] ES cells (Okano, M. *et al., Cell* 99:247-257 (1999)). To determine whether methylation imprints can be stably maintained, the methylation status of a number of imprinted genes was examined at their DMRs using genomic DNA from late-passage 7aabb cells. As shown in Fig. 22, all examined loci, including the 5' upstream region of *H19*, region 2 of *Igf2r*, the DMR of *Peg1*, and DMR1 of *Snrpn*, became completely demethylated in late-passage 7aabb cells, but not in wild-type (J1), 6aa, or 8bb cells. These observations suggested that Dnmt3a and Dnmt3b not only are involved in *de novo* methylation of imprinted genes in male and female germ cells, but may also play a role in maintaining the methylation imprints in the zygote.

[0299] We then examined whether expression of Dnmt3a/3b proteins in 7aabb cells could restore methylation imprints. The 5' upstream region of *H19*, which includes the DMR that regulates expression of *Igf2* and *H19*, is methylated when it is inherited from the father, but unmethylated when it is inherited from the mother. Digestion with the methylation-sensitive enzyme HhaI resulted in a fully methylated paternal band and several weaker undermethylated smaller bands from the maternal allele in wild type (J1) ES cells. Demethylation of this region in 7aabb cells resulted in several lower-molecular-weight bands. We found that Dnmt3a2 almost fully re-methylated this region, whereas Dnmt3a and Dnmt3b1 caused only minimal re-methylation, and Dnmt3b3 and Dnmt3b1:PC showed no activity at all (Fig. 22A). Using similar strategies, we examined several other imprinted genes. DMR2 of *Igf2*, another paternally methylated region, was fully

or partially re-methylated by Dnmt3a, Dnmt3a2, or Dnmt3b1, but not by Dnmt3b3 or Dnmt3b1:PC (Fig. 22B). The intensity of the methylated and unmethylated bands suggested that one allele (presumably the paternal allele) was re-methylated and the other allele remained unmethylated, although we could not rule out the possibility that the methylated band resulted from partial methylation of both alleles. In contrast to *H19* and *Igf2*, none of the maternally methylated genes (*Igf2r*, *Peg1*, and *Snrpn*) could be re-methylated at their DMRs by overexpression of Dnmt3a/3b proteins (Fig. 22C-E). These observations indicate that the maternal methylation imprints, once lost, cannot be restored in ES cells.

Dnmt3b3 inhibits de novo methylation by Dnmt3a and Dnmt3b enzymes

[0300] Consistent with previous results from in vitro DNA methyltransferase assays (Aoki, A. *et al., Nucleic Acids Res* 29:3506-3512 (2001); Okano, M. *et al., Nat. Genet.* 19:219-220 (1998)), our rescue experiments showed that Dnmt3b3 had no enzymatic activity. It is believed that Dnmt3b4, Dnmt3b5, and Dnmt3b6 are also enzymatically inactive because, like Dnmt3b3, they all lack part of the conserved motif IX, due to alternative splicing of exons 21 and 22 (Fig. 20A). To determine whether these isoforms have any activity in vivo, we deleted exons 21 and 22 from the wild-type allele in Dnmt3b+/- and [Dnmt3a-/-, Dnmt3b+/-] ES cells (Okano, M. *et al., Cell* 99:247-257 (1999)) by gene targeting. A PGK-puromycin (PGK-puro) cassette was inserted in the opposite orientation of Dnmt3b transcription to avoid truncation of the Dnmt3b transcripts (Fig. 23A). Since the major Dnmt3b isoforms expressed in ES cells are Dnmt3b1 and Dnmt3b6 (Chen, T. *et al., J Biol Chem* 277:38746-38754 (2002)), we expected that removal of exons 21 and 22 would eliminate Dnmt3b1, but not Dnmt3b6. A number of clones with deletion of the wild-type allele were obtained from both Dnmt3b+/- and [Dnmt3a-/-, Dnmt3b+/-] cells and these clones were referred to as Dnmt3b1KO/- and [Dnmt3a-/-, Dnmt3b1KO/-], respectively (Fig. 23B). Immunoblotting analysis confirmed that Dnmt3b1 protein was abolished and,

concomitantly, the level of Dnmt3b6 protein increased in these cells (Fig. 23C). We examined the methylation status of various repetitive sequences and unique genes in these cells. Unlike the parental Dnmt3b+/- cell line, Dnmt3b1KO/- cells showed significant demethylation of the minor satellite repeats and the methylation pattern was identical to that in Dnmt3b-/- cells (Fig. 23E). Similarly, all sequences examined showed substantial loss of methylation in [Dnmt3a-/-, 3b1KO/-] cells and exhibited methylation patterns indistinguishable from those observed in [Dnmt3a-/-, Dnmt3b-/-] cells (Fig. 23D-E, and data not shown). In addition, [Dnmt3a-/-, Dnmt3b1KO/-] cells failed to methylate newly integrated proviral DNA after infection with a recombinant retrovirus, MoMuLV^{sup}-1, while the parental [Dnmt3a-/-, Dnmt3b+/-] cell line showed efficient de novo methylation activity (data not shown). These data provide genetic evidence that exons 21 and 22 are essential for Dnmt3b activity. We conclude that all Dnmt3b isoforms that lack motif IX have no methyltransferase activity in vivo.

[0301] Interestingly, Dnmt3b3 is ubiquitously expressed and often represents the major Dnmt3b isoform in somatic tissues (Beaulieu, N. *et al., J Biol Chem* 277:28176-28181 (2002); Chen, T. *et al., J Biol Chem* 277:38746-38754 (2002); Robertson, K.D. *et al., Nucleic Acids Res* 27:2291-2298 (1999)). To determine whether Dnmt3b3 plays a regulatory role in DNA methylation, we generated 7aabb-derived cell lines that expressed the active Dnmt3a and Dnmt3b isoforms in the presence or absence of Dnmt3b3. As shown in Fig. 24A, the clones we chose to analyze expressed similar levels of Dnmt3a, Dnmt3a2, or Dnmt3b1. Analysis of a number of sequences revealed that the cell lines co-expressing Dnmt3b3 and Dnmt3a, Dnmt3a2, or Dnmt3b1 consistently showed lower methylation levels than their counterparts expressing the corresponding active isoform alone (Fig. 24B). These results suggest that Dnmt3b3 functions as a negative regulator for de novo methylation.

Dnmt3a/3b-induced remethylation rescues the capacity of [Dnmt3a-/-, Dnmt3b-/-] ES cells to form teratomas in nude mice

[0302] It has been reported that Dnmt1 null ES cells die upon induction of differentiation and cannot form teratomas (Lei, H. *et al., Development* 122:3195-3205 (1996); Tucker, K.L. *et al., Proc. Natl. Acad. Sci USA* 93:12920-5 (1996)). It is not known, however, whether the differentiation defects are caused by loss of methylation or lack of Dnmt1 protein. Unlike Dnmt1 null cells, which lose methylation very quickly, [Dnmt3a-/-, Dnmt3b-/-] ES cells show gradual demethylation during the course of continuous passage, which makes it possible to address the relationship between genomic methylation and cellular differentiation. We injected early-passage (P10) and late-passage (P70) 7aabb cells into nude mice and tested their ability to induce teratomas. While late-passage cells failed to form palpable teratomas (0/3) within 4 weeks, early-passage cells retained the ability to induce teratomas (2/3) despite their much smaller size as compared to those induced by wild type J1 cells (3/3) (Fig. 25A-B). These results indicated that the ability of ES cells to induce teratomas is dependent on the level of genomic methylation, but not the presence of Dnmt3a and Dnmt3b proteins.

[0303] We then asked whether expression of Dnmt3a/3b proteins in late-passage 7aabb cells could rescue the capacity of these cells to induce teratomas. Consistent with their methylation level, stable lines expressing Dnmt3a (3/4), Dnmt3a2 (4/4), or Dnmt3b1 (4/4) were able to induce teratomas in nude mice, whereas those expressing Dnmt3b3 (0/4) or Dnmt3b1:PC (0/4) were not (Fig. 25A). Although the teratomas induced by these stable lines did not reach the size of those induced by J1 cells (presumably because expression of any one isoform could not fully restore the methylation level), histological analysis revealed that all these teratomas contained multiple differentiated cell types (epithelial tissue, cartilage, muscle, etc.) with no obvious differences (Fig. 25B).

Overexpression of Dnmt1 fails to restore global DNA methylation in the absence of Dnmt3a and Dnmt3b

[0304] It has been recently reported that overexpression of Dnmt1 in ES cells results in genomic hypermethylation (Biniszkiewicz, D. *et al., Mol Cell Biol* 22:2124-2135. (2002) To determine whether Dnmt1 could induce de novo methylation in the absence of Dnmt3a and Dnmt3b, we overexpressed Dnmt1 in late-passage 7aabb cells and, as a control, in Dnmt1 null (c/c) ES cells (Fig. 26A). As shown in Fig. 26B and 26C, introduction of Dnmt1 back into Dnmt1 null cells significantly restored methylation of all repetitive sequences and single copy genes examined except for the maternally imprinted gene *Igf2r*, consistent with a previous study (Biniszkiewicz, D. *et al., Mol Cell Biol* 22:2124-2135 (2002). However, overexpression of Dnmt1 in 7aabb cells had little effect on global methylation as compared to the parental cell line, although a slight increase in methylation of repetitive sequences and in the 5' region of *H19* was observed. Likewise, overexpression of Dnmt3a in Dnmt1 null cells could not restore methylation of repetitive elements and unique loci to high levels. These data provide strong evidence that Dnmt1 alone is not capable of methylating genomic DNA de novo and both Dnmt1 and Dnmt3 families of methyltransferases are required for stable maintenance of normal methylation patterns.

Discussion

[0305] Maintenance methylation is a key process that ensures stable inheritance of tissues-specific DNA methylation patterns from cell to cell. It was previously thought that Dnmt1 is solely responsible for the maintenance of DNA methylation patterns since Dnmt1 is expressed ubiquitously and inactivation of *Dnmt1* by gene targeting in mice results in genome-wide loss of methylation (Lei, H. *et al., Development* 122:3195-3205 (1996); Li, E. *et al., Cell* 69:915-926 (1992)). However, there is no evidence that Dnmt1 alone is sufficient to maintain all methylation in the genome. In contrast, our initial studies of embryonic stem cells

lacking the Dnmt3 family methyltransferases suggest that maintenance of methylation of some sequences such as the DMR2 region of *Igf2* and the 5' region of *Xist* requires both Dnmt1 and Dnmt3a/3b (Okano, M. *et al., Cell* 99:247-257 (1999)). In this study, we extended our findings and showed that these enzymes are involved in maintaining global DNA methylation patterns. We demonstrated that inactivation of Dnmt3a and Dnmt3b in ES cells resulted in progressive demethylation of all sequences examined, including repetitive elements, imprinted genes, and non-imprinted genes. These results indicate that Dnmt1 alone is not sufficient for stable inheritance of DNA methylation patterns in ES cells.

[0306] We propose that Dnmt1 is the major maintenance methyltransferase which, in association with the DNA replication machinery, methylates hemi-methylated CpG sites with high efficiency but not absolute accuracy, while Dnmt3a and Dnmt3b, via their de novo methylation activity, function as "proof-readers" to fill in the gaps of the hemi-methylated CpG sites left over by Dnmt1. Consistent with this model is the observation that Dnmt1-/- and [Dnmt3a-/-, Dnmt3b-/-] ES cells exhibit very different kinetics of demethylation. Complete inactivation of Dnmt1 resulted in a 90% reduction of total methyl CpG in the genome immediately after Dnmt1-/- cell lines were established (at 10^6 cells or the first passage) (Lei, H. *et al., Development* 122:3195-3205 (1996)). In contrast, inactivation of Dnmt3a and Dnmt3b resulted in gradual loss of methylation in most genomic sequences and it took more than 70 passages to reach a 90% reduction of global methylation.

[0307] In this study, we demonstrated that both Dnmt1 and Dnmt3 families of methylatransferases are required for stable maintenance of global methylation patterns in mouse ES cells. Our observation that neither overexpression of Dnmt1 in [*Dnmt3a-/-, Dnmt3b-/-*] cells nor overexpression of Dnmt3a in *Dnmt1-/-* cells could restore methylation to normal levels suggests that these two types of enzymes have distinct and non-redundant functions and they act cooperatively to

maintain hypermethylation of the genome. It also confirms that Dnmt1 has little or no de novo methylation activity in vivo.

[0308] Since the Dnmt1 and Dnmt3 families of methyltransferases do not appear to have any sequence specificity beyond CpG dinucleotides (Dodge, J. *et al., Gene* 289:41-48 (2002); Okano, M. *et al., Nat Genet* 19:219-220 (1998); Yoder, J.A. *et al., J Mol Biol* 270:385-395 (1997)), several chromatin-based mechanisms have been proposed to explain how DNA methyltransferases may find their targets in the genome (Bird, A. *Genes Dev* 16:6-21 (2002)). One explanation is that chromosomal regions are not equally accessible to DNA methyltransferases. Consistent with this notion, recent studies of two SNF2 family helicases, ATRX and Lsh, have shown that proteins with chromatin remodeling and DNA helicase activities can modulate DNA methylation in mammalian cells (Dennis, K. *et al., Genes Dev.* 15:2940-2944 (2001); Gibbons, R.J. *et al., Nat. Genet.* 24:368-371 (2000). Similarly, the SNF2-like protein DDM1 has been shown to be essential for methylation of both CpG and CpNpG sites in the plant *Arabidopsis thaliana* (Jeddeloh, J. A. *et al., Nat. Genet.* 22:94-97 (1999)). Another explanation is that accessory factors (proteins, RNA, etc.) recruit DNA methyltransferases to specific genomic sequences or chromatin structures. A number of proteins, including PCNA, DMAP1, HDAC1, HDAC2, pRB, have been shown to interact with Dnmt1 and may recruit Dnmt1 to highly methylated heterochromatin during the late S phase (Robertson, K.D. and Wolffe. A.P. *Nat Rev Genet* 1:11-19 (2000)). The PML-RAR fusion protein and Dnmt3L have been shown to interact with Dnmt3a or Dnmt3b and may recruit these enzymes to RAR response elements and imprinted genes, respectively (Di Croce, L. *et al., Science* 295:1079-1082 (2002); Hata, K. *et al., Development* 129:1983-1993 (2002)). In this study, we provide the first evidence that DNA methylation patterns could also be regulated by expressing different isoforms of Dnmt3a and Dnmt3b. We showed that various Dnmt3a and Dnmt3b isoforms appear to have both shared and preferred DNA targets during the process of re-establishing DNA methylation patterns in highly demethylated [Dnmt3a-/-, Dnmt3b-/-] mutant ES cells. Dnmt3a, Dnmt3a2, and

Dnmt3b1 exhibited substantial activity toward all the repetitive sequences examined but they clearly had sequence preferences, with Dnmt3b1 significantly more potent than Dnmt3a proteins in methylating minor satellite repeats. These enzymes also showed notable differences in methylating certain unique genes. Dnmt3a and Dnmt3a2 were able to methylate the 5' region of *Xist* but Dnmt3b1 was not. Similarly, Dnmt3a2 almost fully restored the methylation status of the 5' region of *H19* whereas Dnmt3a and Dnmt3b1 showed little effect. Given that Dnmt3a and Dnmt3b isoforms show distinct cellular localization patterns (Bachman, K.E. *et al., J Biol Chem* 276:32282-32287 (2001); Chen, T. *et al., J Biol Chem* 277:38746-38754 (2002)), their preferences for different genomic sequences may reflect their differences in chromatin accessibility. It is also conceivable that other factors may interact with various Dnmt3a and Dnmt3b isoforms and target them to different genomic regions. It should be noted that the target specificity of different isoforms was determined by overexpression of each isoform in ES cells, although the results are largely consistent with those obtained from Dnmt3a-/- or Dnmt3b-/- single mutant cells. Genetic studies by inactivating specific isoforms in mice will be necessary to confirm their specificity in development.

[0309] Previous studies have shown that Dnmt3b3 does not have methyltransferase activity *in vitro* (Aoki, A. *et al., Nucleic Acids Res.* 29:3506-3512 (2001)). We now confirm that Dnmt3b3, as well as Dnmt3b6, lacks enzymatic activity to chromosomal DNA *in vivo*. However, these "inactive" isoforms may play an important role in determining the overall methylation level because our co-transfection experiments indicate that Dnmt3b3 may function as a negative regulator for de novo methylation by Dnmt3a and Dnmt3b enzymes. This observation is of potential relevance for understanding regulation of DNA methylation in normal and tumor cells. During development, both the overall level of Dnmt3a/3b proteins and the ratio between different isoforms show dynamic changes. In early embryos, Dnmt3a and Dnmt3b are highly expressed and the major isoforms are Dnmt3a2 and Dnmt3b1, respectively. In most somatic

tissues, Dnmt3a and Dnmt3b are expressed at low levels and the only detectable isoforms are usually Dnmt3a and Dnmt3b3 (Chen, T. *et al., J Biol Chem* 277:38746-38754 (2002)). Our data is suggest that Dnmt3a2 and Dnmt3b1 carry out de novo methylation in early postimplantation embryos to establish the initial methylation pattern, and Dnmt3a, in cooperation with Dnmt1, is involved in maintaining tissue-specific methylation patterns. Dnmt3b3 may play a role in preventing Dnmt3a from methylating CpG islands de novo in normal tissues. Generally, the overall level of DNA methylation is lower in cancer cells than in normal cells and hypomethylation has been correlated with elevated mutation rates and thus may contribute to tumorigenesis (Chen, R.Z. *et al., Nature* 395:89-93 (1998)). However, the cause of hypomethylation in cancer cells is not clear. Dnmt3b3 is overexpressed and often represents the only detectable Dnmt3b isoform in many types of human cancer and cancer cell lines (Beaulieu, N. *et al., J Biol Chem* 277:28176-81 (2002); Chen, T. *et al., J Biol Chem* 277:38746-38754 (2002); Robertson, K.D. *et al., Nucleic Acids Res* 27:2291-2298 (1999)). We propose that overexpression of Dnmt3b3 is a contributing factor for hypomethylation. Other "inactive" Dnmt3b isoforms, such as Dnmt3b4, Dnmt3b5, and Dnmt3b6, may also be overexpressed in certain types of cancers and play a similar role as Dnmt3b3. A recent study has shown that overexpression of Dnmt3b4 may lead to hypomethylation of pericentromeric satellite regions in human hepatocellular carcinoma (Saito, Y. *et al., Proc Natl Acad Sci USA* 99:10060-10065 (2002)).

[0310] Genetic studies have shown that *Dnmt3a* and *Dnmt3b* are involved in the establishment of methylation imprints during gametogenesis (Hata, K. *et al., Development* 129:1983-93 (2002)). Our finding that late-passage 7aabb cells show complete loss of methylation of DMRs of imprinted genes suggests that these enzymes may also play a role in the maintenance of imprinted methylation patterns during embryogenesis. Compared to repetitive sequences, imprinted genes were more resistant to demethylation caused by inactivation of Dnmt3a and Dnmt3b (data not shown). It is possible that maintenance methylation by Dnmt1

is more accurate for single-copy genes than for repetitive elements. While the paternally imprinted *H19* and *Igf2* genes are susceptible to re-methylation by ectopically expressed Dnmt1 or Dnmt3 proteins in mutant ES cells, maternally imprinted genes are completely resistant to re-methylation. We speculate that some essential factors required for the establishment of maternal imprints are present in female germ cells but not in ES cells.

[0311] An interesting observation is that early-passage [Dnmt3a-/-, Dnmt3b-/-] ES cells, which still contain significant levels of DNA methylation, are capable of inducing teratomas in nude mice, whereas late-passage cells, which are more extensively demethylated, completely lose this capacity. This clearly indicates that the presence of Dnmt3a and Dnmt3b methyltransferases (thus de novo methylation activity) is not required for ES cell differentiation and subsequent cellular proliferation. Rather, these processes are dependent on the level of DNA methylation. In keeping with this notion, expression of enzymatically active Dnmt3 proteins (Dnmt3a, Dnmt3a2, and Dnmt3b1), but not inactive forms (Dnmt3b3 and Dnmt3b1:PC), rescued the capacity of late-passage mutant cells to form teratomas. Our results are consistent with previous studies showing that Dnmt1 mutant ES cells undergo apoptosis upon differentiation (Lei, H. *et al., Development* 122:3195-3205 (1996); Tucker, K.L. *et al., Proc. Natl. Acad. Sci. USA* 93:12920-12925 (1996)). Failure to differentiate and proliferate may account, at least in part, for the early embryonic lethality observed in Dnmt1 and Dnmt3 null mutant embryos. A threshold level of DNA methylation may be required for some essential developmental processes. Interestingly, a recent study showed that inactivation of *Lsh*, a member of the *SNF2/helicase* family, results in extensive global demethylation in E13.5 mutant embryos but not embryonic lethality (Dennis, K. *et al., Genes Dev* 15:2940-2944 (2001)). It is possible that embryonic methylation patterns are properly established in Lsh-/- embryos during early development. Further studies are necessary to determine how DNA methylation regulates cell proliferation and differentiation.

[0312] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, this invention is not limited to the particular embodiments disclosed, but is intended to cover all changes and modifications that are within the spirit and scope of the invention as defined by the appended claims.

[0313] All publications and patents mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patents are herein incorporated by reference to the same extent as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.